

Can one identify karst conduit networks geometry and properties from hydraulic and tracer test data?

Andrea Borghi^{c,*}, Philippe Renard^a, Fabien Cornaton^b

^a University of Neuchâtel, Center For Hydrogeology and Geothermics (CHYN), 11 Rue Emile Argand, Neuchâtel 2000, Switzerland

^b DHI-WASY GmbH, Waltersdorfer Strasse 105, Berlin 12526, Germany

^c Gocad Research Group, Laboratoire GeoRessources, Université de Lorraine, Site de Brabois, 2 Rue du Doyen Marcel Roubault TSA 70605, Vandoeuvre-Lès-Nancy FR-54518, France

A B S T R A C T

Karst aquifers are characterized by extreme heterogeneity due to the presence of karst conduits embedded in a fractured matrix having a much lower hydraulic conductivity. The resulting contrast in the physical properties of the system implies that the system reacts very rapidly to some changes in the boundary conditions and that numerical models are extremely sensitive to small modifications in properties or positions of the conduits. Furthermore, one major issue in all those models is that the location and size of the conduits is generally unknown. For all those reasons, estimating karst network geometry and their properties by solving an inverse problem is a particularly difficult problem.

In this paper, two numerical experiments are described. In the first one, 18,000 flow and transport simulations have been computed and used in a systematic manner to assess statistically if one can retrieve the parameters of a model (geometry and radius of the conduits, hydraulic conductivity of the conduits) from head and tracer data. When two tracer test data sets are available, the solution of the inverse problems indicate with high certainty that there are indeed two conduits and not more. The radius of the conduits are usually well identified but not the properties of the matrix. If more conduits are present in the system, but only two tracer test data sets are available, the inverse problem is still able to identify the true solution as the most probable but it also indicates that the data are insufficient to conclude with high certainty.

In the second experiment, a more complex model (including non linear flow equations in conduits) is considered. In this example, gradient-based optimization techniques are proved to be efficient for estimating the radius of the conduits and the hydraulic conductivity of the matrix in a promising and efficient manner.

These results suggest that, despite the numerical difficulties, inverse methods should be used to constrain numerical models of karstic systems using flow and transport data. They also suggest that a pragmatic approach for these complex systems could be to generate a large set of karst conduit network realizations using a pseudo-genetic approach such as SKS, and for each karst realization, flow and transport parameters could be optimized using a gradient-based search such as the one implemented in PEST.

Keywords:

Pseudo-genetic
Stochastic modeling
Karst conduits modeling
Inverse problem
Finite element simulation

Introduction

Karst aquifers are extremely heterogeneous and difficult to characterize [2]. Their heterogeneity is induced by the presence of highly permeable preferential flow paths created by the dissolution of the surrounding rock. Those preferential flow paths are often fractures and bedding planes that are enlarged by dissolu-

tion, often resulting into karstic conduits which are organized in hierarchical networks.

Annable [1] gives an exhaustive overview of the evolution of the conceptual models of speleogenesis over the last two centuries. The conceptual model that is considered in the present study [[63], e.g.] is the following: the karst aquifer is composed by 2 main hydrofacies: the matrix which represents more than 90% of the volume of the aquifer, and has an important storage role; and the conduits which represent a very small volume, but have a very high importance for flow, since they are considered to be responsible for more or less 90% of the total flow.

* Corresponding author. Tel.: +41584690546.

E-mail address: andreafrancesco.borghini@gmail.com (A. Borghi).

Although karst aquifers have already been studied for more than a century they are still very difficult to model [1]. Over the last 3 decades, several numerical modeling techniques have been developed to provide better understanding of these aquifers, but also to manage water quantity and quality. A review of these techniques is provided by Ghasemizadeh et al. [23]. A striking feature of this review is that only direct approaches, i.e. modeling flow and transport when the geometry and the properties of the conduits are known, are described. Following the same line of thought, Saller et al. [52] show the benefits of coupling conduit flow (pipes) with matrix elements, but points out the uncertainty regarding the location of the conduits (which strongly influences the flow field) and the extreme difficulty of calibrating such models.

More generally, among the approaches used to solve inverse problems [57] in groundwater hydrology, gradient-based methods are frequently used. They consist in modifying iteratively the hydraulic conductivity values either in predefined zones [10] until the error between observed and calculated state variables reaches a minimum or stabilizes at an asymptotic value. This is extremely efficient if zones of constant but unknown hydraulic conductivity values are predefined. If the spatial distribution of the permeability field itself is unknown, it can be inferred as well using techniques such as pilot points, gradual deformation, or probability perturbation methods [9,30,48] which are also using gradient optimization.

In the case of karstic aquifers, the progressive deformation of the conduit structure and especially topology to reach a maximum likelihood or minimal error solution is particularly difficult to achieve. Preliminary tests conducted in this research have shown that the application of methods such as the probability perturbation lead to extremely discontinuous behavior of the objective function, making the use of gradient optimization completely useless. This is explained by the fact that during the progressive deformation of the geometry, disconnection and reconnection of the conduits occur leading to abrupt changes in the hydraulic and transport response.

Such difficulties explain partly why there are so few studies that considered applying inverse methods for karst aquifer distributed parameter models. Notable exceptions are Larocque et al. [40,41] and Panagopoulos [46] who used inverse methods to calibrate karst hydrogeological models. But, both group of authors do not include discrete karst conduits in their models and represent the karst aquifer with a 2D equivalent porous medium. Moreover, in these previous works, transport data have not been considered.

In this perspective and before conducting or developing any inverse methodology, it is important to understand better which parameters (like the hydraulic conductivity and porosity, or the shape and the number of conduits) mainly control the simulation results and therefore the value of the objective function.

Moreover, the ability of a classical optimization technique to effectively (and possibly efficiently) calibrate the physical properties of karst aquifers has also to be tested when simulating complex systems, with Darcy laminar flow in the matrix and turbulent Manning–Strickler flow in discrete pipes.

In this perspective, the present paper investigates whether inverse methods could be used to obtain information about the structure of the karstic network or the distribution of the conduit dimensions.

Two distinct numerical experiments are carried out. The first considers the influence of changing geometry and topology of the karst conduits on the simulation results. 150 different geometries with varying karst conduit radius and matrix hydraulic conductivity are considered. All these models are generated using the pseudo-genetic algorithm previously developed by Borghi et al. [8]. In practice, the test is performed by comparing the results of 18,000 2D flow and transport EPM (Equivalent Porous Medium) finite elements simulations (Sections 3). The second test investigates

the ability of an inverse algorithm such as PEST [17] to retrieve optimal physical parameters when using a more complex model, i.e. a 3D model with karst conduits meshed as 1D pipe elements and non linear flow dynamics in the conduits.

1. Literature review

1.1. Flow simulation

Mathematical black-box models consider the whole aquifer as a single reservoir whose global behavior can be described with simple mathematical relations between an input signal and an output response: e.g. global parameter models [e.g. [42]], or neural networks [e.g. [29]]. An exhaustive review of these kind of models can be found in Ghasemizadeh et al. [23]. Unfortunately, as explained by De Marsily [14], these models may be sufficient to predict spring hydrographs, but they do not provide any spatial information about the karstic conduit system.

As opposed to black-box models, distributed parameter models are based on the discretization of the model domain into sub-units. Each sub-unit has homogeneous parameters in the space that it delimits (see Section 2.2). As Ghasemizadeh et al. [23] say, *the challenge of distributed modeling approaches to represent karst groundwater systems is to cope with the high spatial heterogeneity of karst aquifers*. Many authors have already modeled karst aquifers using parameter distributed models. The easiest way is to consider the whole karst aquifer as an equivalent porous medium (EPM), where matrix, fractures and conduits are brought together in an equivalent hydrofacies (e.g. [6,39]). Unfortunately, EPM models have a very low applicability in very karstified fields and may lead to catastrophic situations like the case of Walkerton (Ontario, Canada) where, in May 2000, 7 people died from a bacterial contamination of the municipal water supply because the spring protection zone was based on an EPM model, which gave much larger transit time than what was observed (later) by field tracer tests (Worthington, [64]; Goldscheider, [25]; Kresic and Stevanovic [37]).

To avoid this kind of issues, other authors have developed models, which use the available information about the conduit geometry to add heterogeneity in their model. Worthington [62] models the Mammoth Cave aquifer using MODFLOW, and he defines the mesh elements where karst conduits were explored as cells with higher hydraulic conductivity. He also had to increase the hydraulic conductivity according to the hierarchy of the conduits to be able to simulate realistic head distributions. Király [33], Király et al. [35] modeled the conduits as discrete 1D or 2D features embedded in a 3D matrix using a discrete-continuum approach, flow in conduits being laminar. The discrete-continuum approach is often used to develop speleogenesis models [1,18,22]. These speleogenesis models are useful to understand the complex kinetics of karst aquifers. In addition, as Jeannin [32] pointed out, turbulence is often observed in conduit flow. Nowadays there are computer codes that allows non linear flow equations to be used in discrete conduits, as MODFLOW-CFP [Conduit Flow Process, [38]] or GROUNDWATER [12]. De Rooij [15] developed a model that is able to simulate also unsaturated flow in pipes. Recently, Saller et al. [52] show the benefits of the application of a discrete conduit model for the simulation of the Madison aquifer of Southern Dakota (USA).

1.2. Conduit network modeling

The models of De Rooij [15] and Saller et al. [52] show enhanced results with respect to EPM models, because they solve more complex physics, but their authors agree on the difficulty that is posed by the unknown conduit location. In order to use realistic conduits, one could use the networks resulting from speleogenetic models [1,18,22], but the computation of these models is

heavy and they are difficult to condition to field data. Geostatistical based models can be used as an alternative to obtain discrete conduit models like Fournillon et al. [21]. They are easier to condition to field data, but it is difficult to produce well connected paths. Jaquet et al. [31] propose to use a modified lattice-gas automaton for the discrete simulation of karstic networks, which can produce hierarchical structures, but which is also difficult to condition to field data. Ronayne and Gorelick [50] propose to simulate branching channels (analogous to karst conduits) using a nonlooping invasion percolation model, which produces nice features, but that are based on a pixel discretization to represent the enlargement of the conduits, and are therefore very difficult to use directly for a flow simulation. Finally, a new approach is to use a pseudo-genetic algorithm [8,11,28], which mimics the resulting structures of karst networks produced by speleogenetical models, without solving the complete kinetics of the system, i.e. the combination of physical and chemical processes that lead to speleogenesis: equilibrium state of calcite (rate of dissolution/precipitation), transport of carbonates, etc. These computations are extremely computationally expensive. On the other hand, pseudo-genetic models use approximated physics to generate the karst networks, which are extremely lighter to compute. The approach of Borghi et al. [8] is based on the assumption that water (and consequently conduit development) follow a minimum effort path as suggested by Groves and Howard [26]. The conduit enlargement and development does not strictly follow physical and chemical laws during the simulation, but the resulting conduit networks satisfyingly mimic the networks obtained using full speleogenetic models.

1.3. Recharge and epikarst model

Karst aquifers present several complex recharge phenomena, which are caused by the presence of a highly altered superficial "skin" (of a few meters) on the top of the aquifers, called epikarst [20]. The epikarst has a complex role on aquifer recharge, because it concentrates the rainfall into several "point" inlets (dolines), which lead directly to the karstified network. It represents roughly the "skin" of karst aquifers, that rapidly drains and concentrates the rainfall into several infiltration points known as dolines (concentrated recharge). Moreover, it is quite common that surface water streams can infiltrate directly into the conduit networks through other point inlets called sinkholes.

The conceptual model of Mangin [42] considers that a consistent part of the rainfall is quickly drained by the epikarstic layer toward the conduit network, and the remaining part of water will be infiltrated diffusely on the low-permeability fractured volumes. Király et al. [36] use a nested 2D model to simulate the epikarst, coupled with a 3D model where the karstic network was modeled. The epikarst model outputs are used as input for the concentrated recharge (sinkholes and dolines). Their recharge function for the flow model underlying the epikarst layer is computed as a proportion of diffuse and concentrated recharge. Their final conclusions admit that in *most open karst aquifers more than 40% of the infiltration should be drained rapidly into the karst channels*. This work showed the benefits of correctly modeling the effect of epikarst on recharge. Authors like Ofterdinger et al. [45] use a complex hydrological model that computes the recharge contribution and the runoff of precipitation on the base of several parameters, like the slope, the altitude and the soil cover. They obtain satisfactory results. Unfortunately, this kind of model can be very difficult to correctly parameterize, especially if only a few measurement stations are available for model calibration. Finally, Weber et al. [61] use black-box models to estimate the recharge of a karst aquifer. Their approach shows very promising results and they demonstrate that a good estimation of the recharge function is essential to correctly model a realistic flow behavior.

1.4. Inverse problem

The aim of the inverse problem is to identify the geometry, physical parameters, initial or boundary conditions of any model that describes a system from field observations of state variables [51,57]. Unfortunately, due to their mathematical structure, inverse problems are usually ill posed and have either no solution, an infinity of solutions, or can be unstable [27]. Many different approaches have been developed to overcome these difficulties over the last 60 years.

One of the most general way to solve the inverse problem is to frame it into a Bayesian framework. This implies to define a prior probability distribution for the unknown input parameters and a statistical distribution of the acceptable error on the measured state variables. Based on these two main ingredients one can formulate the expression of a likelihood and deduce a posterior probability distribution for the unknown parameters. This formalism is described in detail in [57].

Then, one can either search only for the parameter set leading to the maximum likelihood using optimization techniques or aim to a more complete solution by sampling directly the posterior distribution. Sampling the posterior requires using Monte Carlo techniques [53]. Among those, the most simple is the rejection-sampling method [59]. It consists in generating a large number of candidate models within the prior distribution and accepting or rejecting them with a probability that depends on the probability distribution of the expected measurement errors [57]. The result is a number of models that are proper samples of the posterior probability distribution of the unknown parameters. However, this method is computationally inefficient since it requires running a very large number of models. It is therefore used only for benchmarking more efficient methods or to study particularly simple models as it will be done in the following of this paper.

In practice, solving the inverse problems in general to find a compromise between computational feasibility and proper uncertainty quantification. It also requires to use some adequate tools that one can couple to any forward model. One of the most flexible tools available today for such model parameter estimation is PEST [17]: it offers a wide range of methods that can be coupled to any model. It includes functionalities for the minimization of an objective function using Levenberg-Marquardt method [43,44] as well as Null Space Monte Carlo sampling [58] for example.

2. Karst modeling method

This section describes the forward modeling methods that are used in the two following sections when we investigate the inverse problems.

2.1. Conduit networks

In this paper, the Stochastic Karst Simulator (SKS) pseudo-genetic method developed in Borghi et al. [8] is used to model the 3D geometry of conduit network. The method includes the following steps: (1) a 3D geological model is built; (2) a stochastic fracture model is used to add heterogeneity into the 3D geological model [7]; (3) the conduits of the karstic network are generated using the pseudo-genetic approach, which uses a Fast Marching Algorithm [FMA, [56]] to compute minimum effort paths between karstic inlets (dolines and sinkholes) and springs. This minimum effort path computation is based on the assumption that the water (and consequently the conduits generated by dissolution) will preferentially flow inside the more conductive discontinuities like fractures and bedding planes, termed as inception horizons [19]. Moreover, the conduits are generated iteratively. Every conduit that has already been simulated influences the next ones, because it is

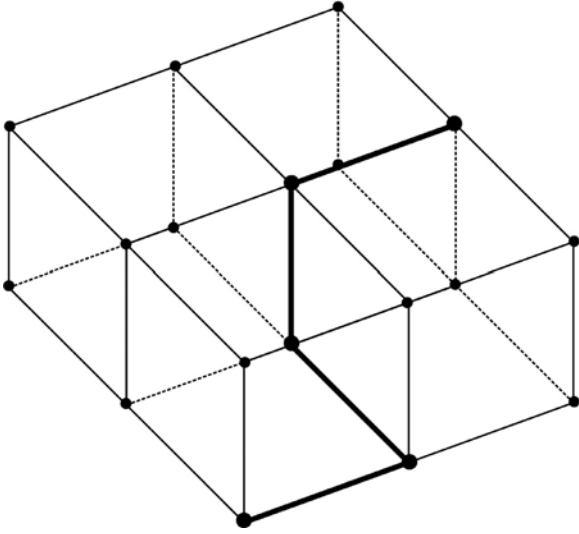


Fig. 1. The conduits are meshed as 1D pipes that follow the edges of the 3D voxels (matrix).

considered as a new preferential flow path. This leads to a hierarchical network, where the conduits that start from several inlet points converge toward the outlets of the system. SKS allows to distinguish the saturated and unsaturated zones, to account for several phases of karstification, and can generate cycles and complex interconnections between the conduits.

2.2. Meshing

In its current implementation, the model is meshed with bricks in 3D and quadrangles in 2D. If an Equivalent Porous Medium approach is used to model the flow (as done in Section 3), the cells corresponding to the conduits are flagged as conduit elements. Otherwise, if a flow simulation based on laminar or turbulent 1D flow equation for discrete conduits must be performed (Section 4), the conduits are meshed as 1D line elements, which are composed by 2 nodes and follow the edges of the 3D bricks. As shown in Fig. 1, the pipe nodes are consistent with the 3D finite element grid, which allows using a discrete continuum approach to simulate the interactions between matrix and pipes.

2.3. Recharge black box model

A reservoir black-box model is used to simulate the total recharge function for the finite element model. Two kinds of recharge patterns are then considered: diffuse and concentrated. Diffuse recharge is directly assigned on topographic surface nodes of the model, and concentrated recharge is assigned to the nodes connected to dolines or sinkholes. The proportion λ_c of the total recharge that will be considered as concentrated is defined by the user and is used to compute both the total concentrated recharge R_c and the total diffuse recharge R_d :

$$R_c = \lambda_c R, \quad R_d = R - R_c \quad (1)$$

where R is the total recharge. Then the concentrated recharge R_c^i for each inlet i is calculated with respect to their estimated catchment surface S_i over the total N dolines catchment surface:

$$R_c^i = R_c \frac{S_i}{\sum_k^N S_k} \quad (2)$$

R is a function that varies with time and that depends on the meteorological conditions. It can be estimated for example using the

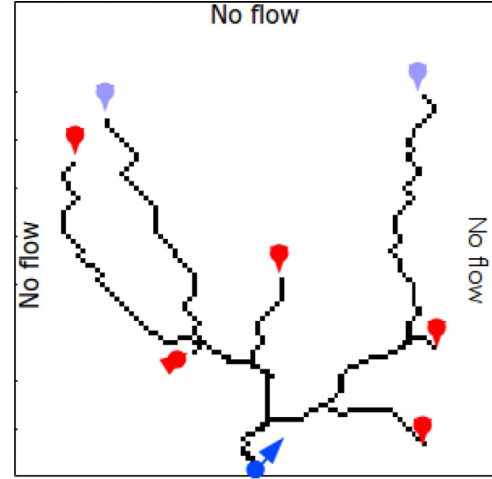


Fig. 2. Model description: in white the matrix elements, in black the conduits; Flow boundary conditions: the blue arrow is the spring (constant head), the inlets have 90% of total recharge, the remaining 10% is distributed on the matrix elements; Transport: injection of 1 kg of tracer in 2 inlets (violet point). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

software RS2012¹, which is a hydrology routing system developed for watershed management.

2.4. Flow and transport boundary and initial conditions

2.4.1. Flow boundary conditions

Spring zones, and in general the discharge areas, are modeled with prescribed head boundary conditions: the outlets of the system are supposed to be at fixed altitudes, and the outflows depend on the head variations only.

On the other hand, two kinds of boundary conditions can be used for recharge:

- Prescribed flux of Neumann type (m/s) at the boundary of an element
- Prescribed inflow (m³/s) assigned to the nodes that correspond to sinkholes and dolines

Neumann fluxes are considered to be perpendicular to the surface. GROUNDWATER uses the net inflow projecting the vertical component of the inflowing flux onto the face of the corresponding element.

A source term can alternatively be used to model recharge. It corresponds to the direct injection of water into the elements at the topographic surface.

In this way, it is possible to infiltrate the corresponding value of recharge into the nodes.

2.4.2. Transport boundary conditions

In the proposed workflow, the tracer tests are simulated separately: every one starts at injection time using an initial concentration of tracer at the injection node. The initial concentration C_{in} is computed from a known mass to be injected as follows:

$$C_{in} = M/V_p \quad (3)$$

where M is the injected mass (kg), and V_p is the porous volume at the given node (i.e. the node of the injection). The porous volume

¹ The reader can find the documentation of the software at <http://e-dric.ch/index.php/en/software-en/rs2012>.

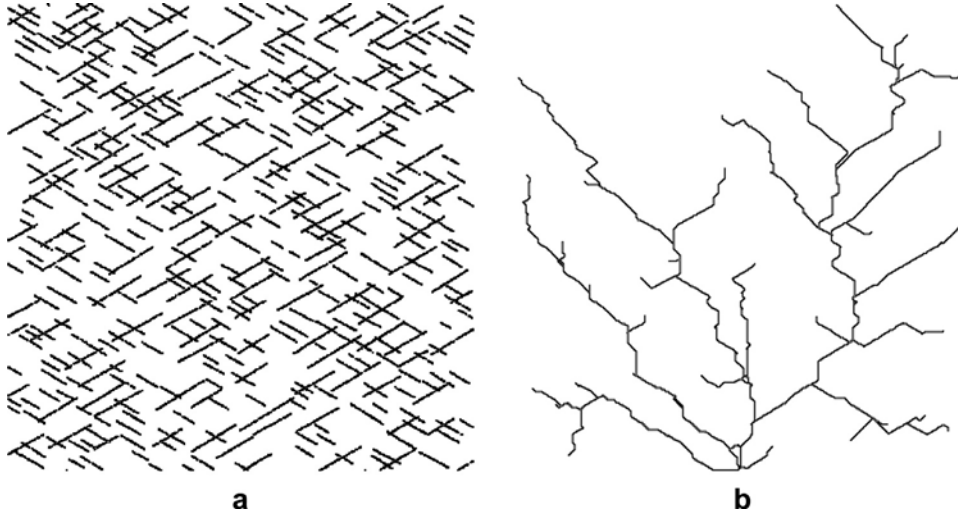


Fig. 3. A stochastic fracture network (a) is the basis for the karst conduits model (b).

of a given node is defined as:

$$V_p = \sum_{i=1}^{N_i} \frac{\phi_i V_e}{8} \quad (4)$$

where i runs on all the parallelepipedic elements that are connected to the given node, V_e is the element volume, ϕ_i is the porosity of each element. If the simulation uses also 1D pipe elements, and that the given node is also connected with a 1D element, Eq. 4 becomes:

$$V_p = \sum_{i=1}^{N_i} \frac{\phi_i V_e}{8} + \sum_{j=1}^{N_j} \frac{\pi r_j^2}{2} \quad (5)$$

where j are the 1D elements connected to the node, and r_j is the radius of these elements.

These equations show that the initial concentration strongly depends on the discretization, on the elements that are connected to the given node, and also on their properties (i.e. porosity, radius). Moreover, if recharge is applied to the injection node, it may additionally dilute the solute. To account for this dilution, the inflowing recharge Q_t at a time t with duration of dt is summed over the porous volume V_p as follows:

$$V_p^t = V_p + \frac{Q_t(t)}{dt} \quad (6)$$

where V_p^t represents the “porous” volume at time t , i.e. the water volume that dilutes the tracer at time t .

3. Retrieving conduit geometry

In the present section, we analyze if geometric or physical properties of a karstic system can be identified using an inverse approach. To answer this question, a numerical experiment is performed in which an ensemble of different geometries and parameter sets are generated. Then we search systematically which other geometries or parameter sets produce similar responses.

3.1. Model description

The case study considered in this numerical experiment is a synthetic 2D finite element model with 200x200 m size and 2 m resolution ($dx=dy$). The flow and transport equations are solved using the finite element code GROUNDWATER [12].

The model considers 2 hydrofacies: “karst” and “matrix” (NB: karst elements are those that contains a karst conduit). Flow in

karst and matrix elements follow Darcy’s law (laminar flow); an EPM² is therefore used within the “karst” elements. The model assumes steady-state flow and transient transport. The synthetic case mimics a small karst aquifer, for which only 2 inlets (sink-holes) and 1 spring are known. For both inlets, the mean head is supposed to be known and a tracer test has been performed. The tracer test is a punctual injection³ of a unitary mass at time 0 of simulation. The flow boundary conditions are: no flow boundary on every side of the model, a fixed head at the spring node and a constant recharge over the model. 90% of the total recharge is directly infiltrated in the inlets of the model and only 10% on the “matrix” elements (i.e. $\lambda_c = 0.9$, in Eq. 1). Leaving 10% of recharge on the other elements allows the reproduction of a gradient from every element of the model toward the spring. Fig. 2 summarizes the model setup.

3.2. Geometrical karst realizations

150 geometrical realizations are generated using the pseudo genetic methodology described in Borghi et al. [8] (Fig. 4). All realizations can be grouped into 3 families based on the number of inlets (50 realizations for every family). The first family has only 2 inlets (i.e. the 2 known inlets), the second 7 (i.e. 2 of known location, and the others with random locations) and the third 14 inlets (of which 2 are deterministic). Note that SKS generates conduits from every inlet toward the spring. It means that a simulation with 2 inlets will correspond to geometries with 2 main conduits. These conduits can merge and create a hierarchical structure that can be seen as a tree with three branches, but for the sake of simplicity, we will describe these karst networks as made of two conduits and similarly for those with 7 or 14 inlets. The pseudo genetic methodology requires the existence of an initial fracture network to control the variability of the realizations of karst networks. Here, a different equiprobable realization of a fracture model is used for every karst realization. The same fracture statistics are used for every realization so that the final karst realizations still remain statistically comparable. Fig. 3 shows one realization of the fracturation model (a) and the corresponding stochastic karst conduits model (b).

² Appendix B details the EPM computation.

³ Appendix D explains the necessary conditions to ensure the numerical stability of the tracer test simulation.

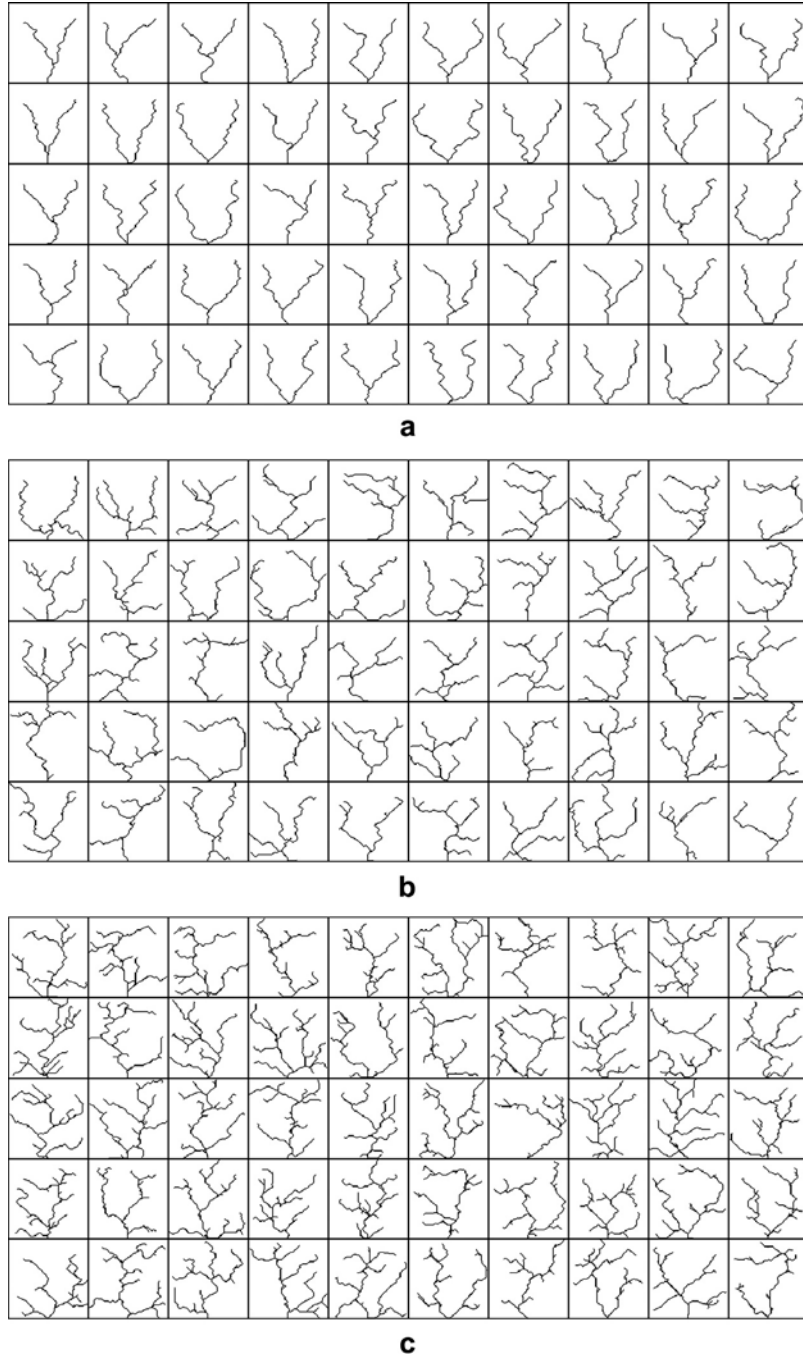


Fig. 4. Stochastic karst realizations used for this study, 50 realizations with (a) 2 conduits, (b) 7 conduits, and (c) 14 conduits.

3.3. Flow parameters for each karst realization

For every karst realization, 120 different combinations of the hydraulic conductivity of the matrix K and conduit radius r were tested. The hydraulic conductivity of the matrix K varies from $5 \cdot 10^{-7}$ m/s to 10^{-3} m/s (8 values with regular steps in a \log_{10} space) and the radius r of the conduits in karst elements, varying from 0.1 to 0.8 m (15 values with an increment of 0.05 m). It is important to note that changing the radius of the conduits leads to differences in both porosity n (Eq. B.1) and equivalent permeability κ_{eq} (Eq. B.4). Fig. 5 shows the variation of both porosity n and equivalent hydraulic conductivity K_{eq} with an increasing radius r from 0.1 m to 0.8 in 2D cells of 2 m length.

3.4. Reference simulation selection

After having solved the direct problem for both tracer tests and for every combination of parameters for every karst realization (which means $150 \times 120 \times 2 = 36,000$ simulations) one reference simulation is chosen. It is selected from the 18,000 flow parameter fields (150×120). This reference simulation is then considered as our reality. The results of this simulation are 2 observations of head H^i at both inlets, and 2 tracer breakthrough curves C^i at the spring.

These observations are then compared to the results of all the other simulations. The comparison is based on the following objective function:

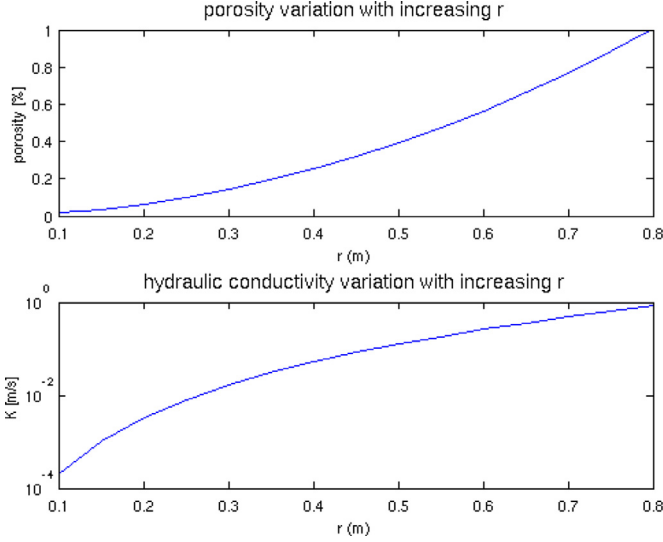


Fig. 5. Variation of the equivalent porosity [-] and hydraulic conductivity [m/s] as a function of the radius r of the conduits .

$$e = \lambda e_c + (1 - \lambda) e_H \quad (7)$$

with:

$$e_c = \frac{\frac{1}{N_c} \sum_{i=1}^{N_c} |C_{obs}^i - C_{calc}^i|}{\max\left(\frac{1}{N_c} \sum_{i=1}^{N_c} |C_{obs}^i - C_{calc}^i|\right)} \quad (8)$$

and

$$e_H = \frac{\frac{1}{N_H} \sum_{i=1}^{N_H} |H_{obs}^i - H_{calc}^i|}{\max\left(\frac{1}{N_H} \sum_{i=1}^{N_H} |H_{obs}^i - H_{calc}^i|\right)} \quad (9)$$

where λ is a weighting percentage, N_c are the number of time measurements of concentration, and N_H the number of head observations.

3.5. Inverse problem

Having a given reference data set, solving the inverse problem consists here in finding the ensemble of configurations (geometry + parameters) that match the observations. In practice, the error e is computed using Eq. 7 for all the other models. Only the one for which e is lower than a given threshold are selected. The result is an ensemble of realizations that match the data and describe statistically the remaining uncertainty.

Fig. 6 illustrates this procedure. In Fig. 6a, the ensemble of all the 18,000 recovery curves are displayed with the one that is considered as the reference. Huge differences between the reference and most of the other breakthrough curves are observed. Fig. 6b shows the selected acceptable simulations, i.e. those displaying an error e (Section 3.4) lower than 1%.

Fig. 7a shows the histogram of e_c and Fig. 7b the histogram of e_H in this case. The response for transport (Fig. 7a) is much more discriminant than the one for flow (Fig. 7b). The histogram of the errors on the flow responses shows that it is quite easy to obtain a good fit, as many simulations have a very small value of e_H . The reason for this behavior is that with only few observations, it is easy to obtain several models that match the observations within a given threshold of confidence. On the contrary, transport results show a wider distribution of the errors e_c , related to a higher sensitivity of the solution to the geometry of the flow paths. Indeed, the error on the breakthrough curves is highly influenced by the path followed by the streamlines. Fig. 7c shows the histogram of the objective function with a value of λ equal to 1/2 (Eq. 7), i.e. giving the same weight to both e_c and e_H .

3.6. Statistical analysis

The previous section has shown that it is possible to find models matching a certain reference. In the present section, the experiment is repeated using systematically all models as a reference. The inverse problem is solved for each one (as explained in Section 3.5). This allows analyzing in a statistical manner if the parameters are well identifiable.

To represent the results, the probability distribution of the estimated parameter values is computed and compared to the true parameter values in the form of log ratios: $\log_{10}(r_{est}/r_{ref})$ and $\log_{10}(K_{est}/K_{ref})$. These probabilities are represented conditional to a certain value of the reference parameters:

$$P(\log_{10}(r_{est}/r_{ref}), \log_{10}(K_{est}/K_{ref}) | N_{ref} = x) \quad (10)$$

where x is the possible number of conduits. When $\log_{10}(r_{est}/r_{ref})$ or $\log_{10}(K_{est}/K_{ref})$ are equal to 0, it means that the estimated radius r_{est} and the estimated matrix conductivity K_{est} are equal to the reference. We also compute the probability of having N_{est} number of estimated conduits, knowing the N_{ref} number of reference conduits:

$$P(N_{est} | N_{ref}) \quad (11)$$

which allows understanding the probability to identify properly the number N of conduits of the reference. We also compute the conditional joint pdf of the ratios: $\log_{10}(r_{est}/r_{ref})$ and $\log_{10}(K_{est}/K_{ref})$ knowing N_{ref} and N_{est} the number of conduits of the possible solutions:

$$P(\log_{10}(r_{est}/r_{ref}), \log_{10}(K_{est}/K_{ref}) | N_{est} = y, N_{ref} = x) \quad (12)$$

where x and y are the possible number of conduits.

Figs. 8 and 9 show the results. They suggest that the conduit radius is likely to be better identified by the inverse procedure than the hydraulic conductivity of the matrix. The probability of having an estimated radius larger or smaller than twice the real radius is very small; a ratio in the range [1/2, 2] between the real and estimated radius corresponds to a \log_{10} value in the range [-0.3, 0.3]. We see that outside of this range, the probabilities are very low.

On the opposite, the probability to have an estimated hydraulic conductivity for the matrix around 3 orders of magnitude larger or smaller than the truth is large. This is consistent with our conceptual model, for which 90% of the flow and possibly all the tracer are supposed to pass through the conduits.

One interesting thing that can be seen on Fig. 8 is that when the reference has only 2 conduits ($N_{ref} = 2$), the probability $P(N_{est} = 2 | N_{ref} = 2)$ is extremely high as compared to $P(N_{est} = 7, 14 | N_{ref} = 2)$. In this case the inverse problem is quite robust and able to identify precisely that the unknown reality has indeed 2 conduits. This result is of course valid only under the tested configuration (2 inlets, 2 conduits, 2 tracer tests).

On the opposite, when the reference has 7 conduits ($N_{ref} = 7$), the probability to estimate that the model should have 7 inlets is only slightly higher than the probability of having 2 or 14 conduits. This is even more difficult when the reference has 14 conduits ($N_{ref} = 14$). This is related to the fact that only 2 tracer tests are available in this synthetic example. But it shows that when sufficient data are available, one can expect that the inverse problem will allow to identify properly some major features of a karstic network (such as the number of main conduits or their radius).

Looking more in detail at Fig. 9, we notice that the estimated radius r_{est} and matrix hydraulic conductivity K_{est} are very well identified for the simulations for which $N_{est} = N_{ref}$ (Fig. 9a, e, and i), i.e. both \log_{10} -ratios are close to 0. When $N_{est} < N_{ref}$, the estimated values of r_{est} and K_{est} are overestimated. This makes sense because the model needs to drain more water relatively to the number of conduits. For the same reason, when $N_{est} > N_{ref}$ the estimated values of r_{est} and K_{est} are underestimated.

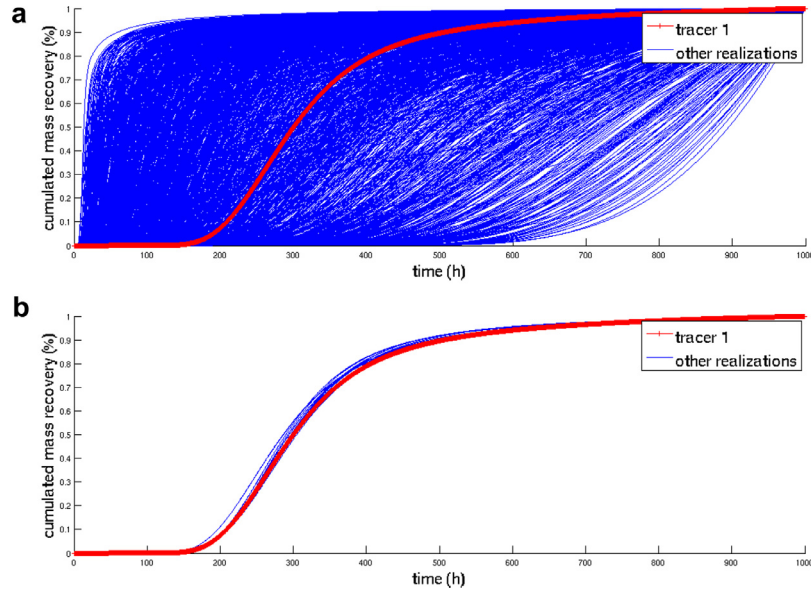


Fig. 6. Tracer test results (cumulated mass at the spring). (a) In red we see the tracer test result for tracer one on the reference simulation, in blue the results of all the other simulations. (b) selection of the “best fit” simulations, i.e. the ones under 1% of error. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

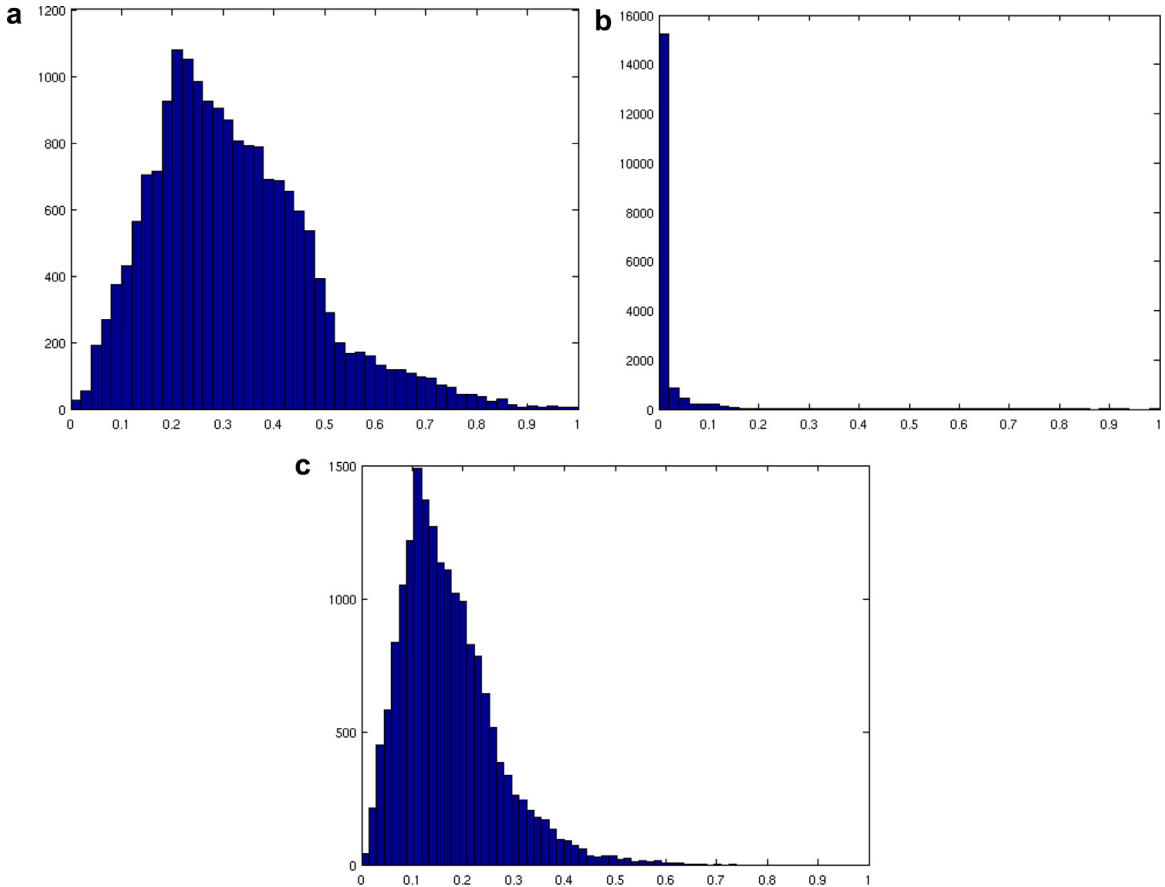


Fig. 7. pdf of normalized error, (a) on tracer test results (e_c). (b) on flow results (e_H). (c) using both flow and transport (e), with $\lambda = 0.5$.

4. Retrieving conduit properties using gradient-based methods

In the previous section, the acceptable parameters are obtained by rejection sampling and systematic search implying to run the model for many possible geometries or parameter values. This was possible because the model was small and simple. For a large scale

model (regional scale, millions of nodes) this is generally not possible and a faster method is required. This is why we test in this section the feasibility to use a gradient based approach to accelerate the identification of the physical parameters when the geometry is fixed. For that purpose, we run a systematic parameter search to know exhaustively the objective function and then test

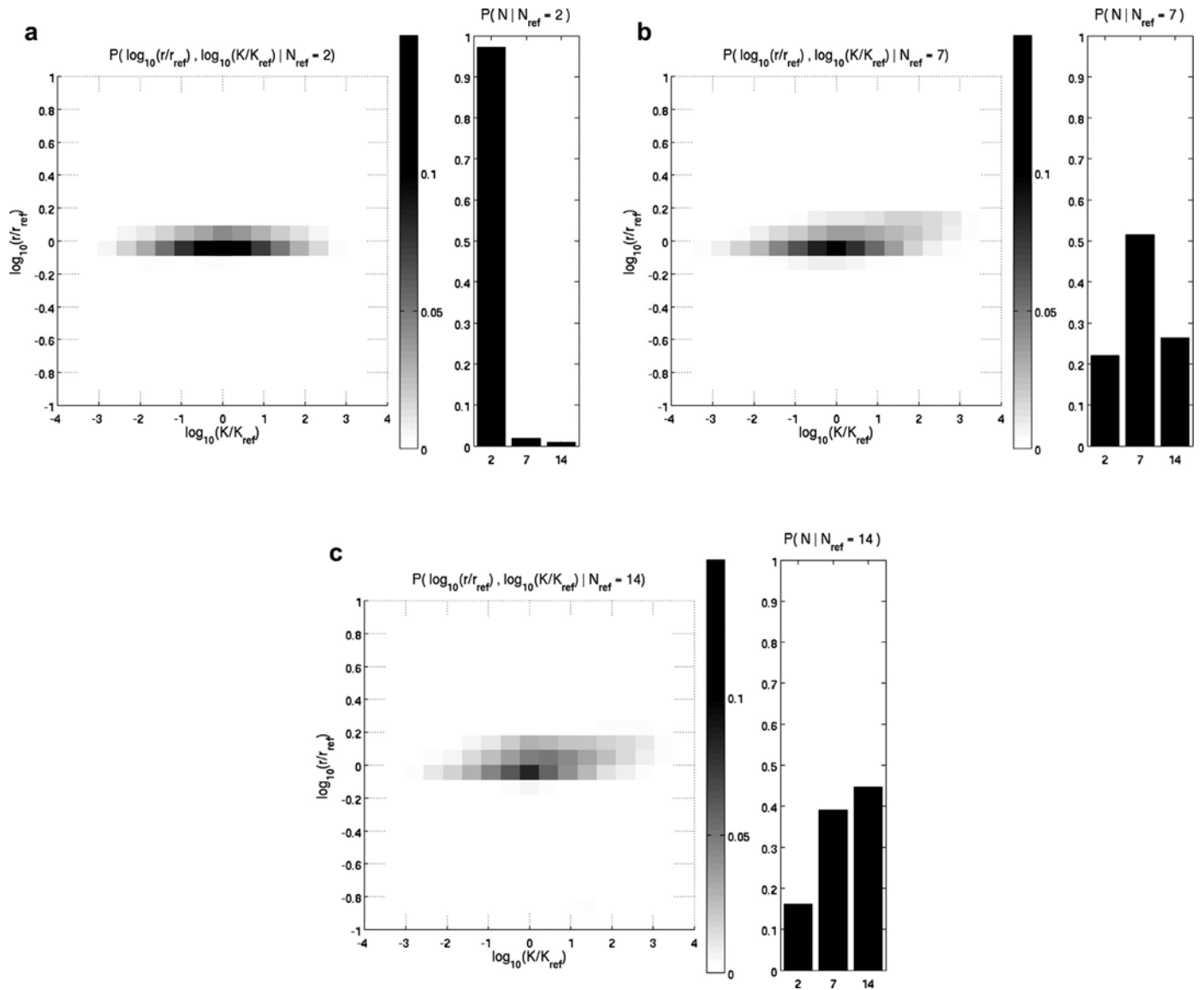


Fig. 8. Joint probability distribution functions obtained using rejection sampling: (a) case with a reference having 2 conduits ($N_{ref} = 2$); (b) case with $N_{ref} = 7$; (c) case with $N_{ref} = 14$.

the efficiency of the gradient search using the software PEST [17] on a synthetic case.

4.1. Model description

Karst networks often show some degree of hierarchy [34]. As Annable [1] demonstrates by speleogenetical modeling, karst conduits can show an organization of the mean conduit diameter that increases according to the hierarchy of the network, i.e. that the biggest conduits are located downstream as they collect the water of their affluent conduits in a similar way as rivers. Differently from streams, karst conduits can also diverge, and in this case the flow is divided between the different affluent conduits [1]. To consider these particular characteristics of karst networks, in a previous work [60] we have proposed a simple model in which the conduit radius r is estimated using a power law similar to Horton's law:

$$r(u) = \alpha e^{u\beta} \quad (13)$$

where α and β are two parameters that have to be calibrated, and u is the order of the conduit. However, unlike Collon-Drouaillet

et al. [11] and Vuilleumier et al. [60] who used a Horton's order based on the river Strahler classification, here the order u of each conduit is computed as the ratio of the catchment surface S that is drained by the given conduit over the total catchment surface as suggested in Borghi et al. [8]:

$$u_i = \frac{\sum_{j \in A_i} S_j}{\sum_{k=1}^N S_k} \quad (14)$$

where N is the total number of inlets of the system, S_k represents the total surface of the catchment that is drained by karst conduits, and A_i represents the set of indices that drain into the conduit i (Figs. 10 and 11).

SKS uses walkers to compute the paths of the conduits. To consider the diverging conduits, it stores the information of the origin of the walkers for each computed path. In this way, it is possible to keep the divergence of conduits consistent with their radius. In the proposed approach, the idea is to create the minimal hydraulic necessary dimensions to allow the system to drain recharge water out of the system.

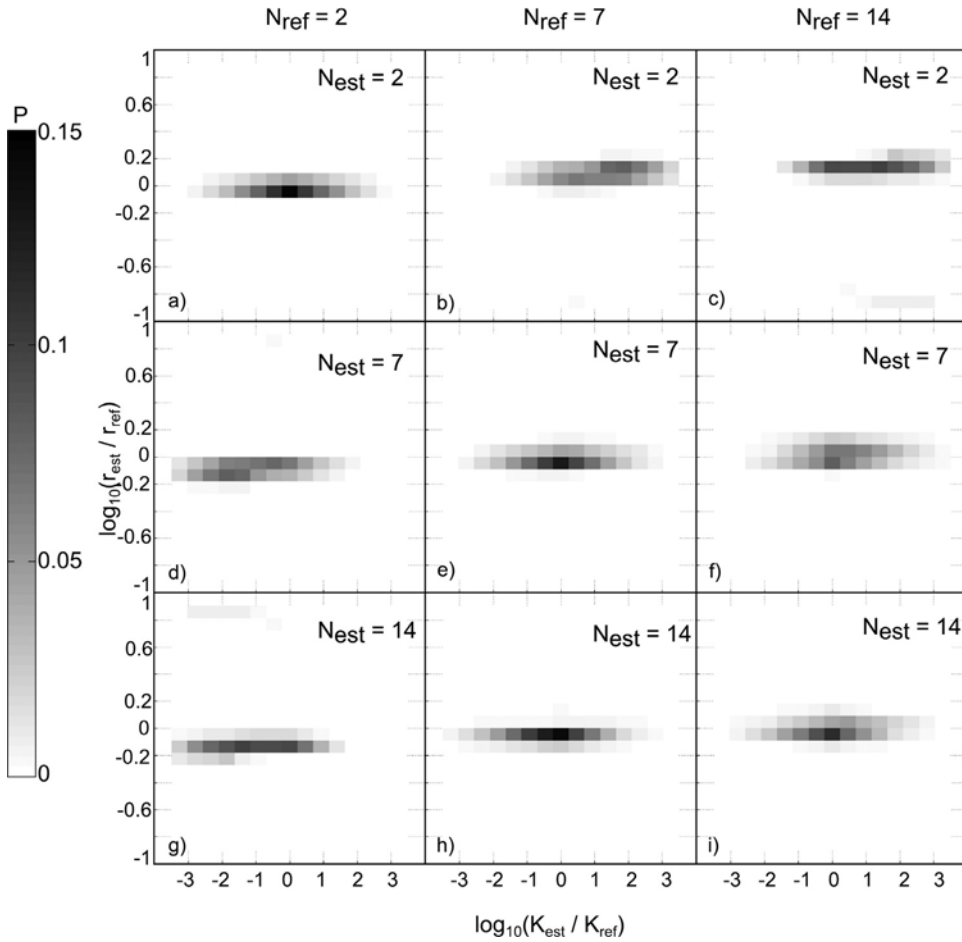


Fig. 9. Joint probability distribution functions of the estimated conduit radius r_{est} and matrix hydraulic conductivity K_{est} as a function of the estimated N_{est} and reference N_{ref} number of conduits.

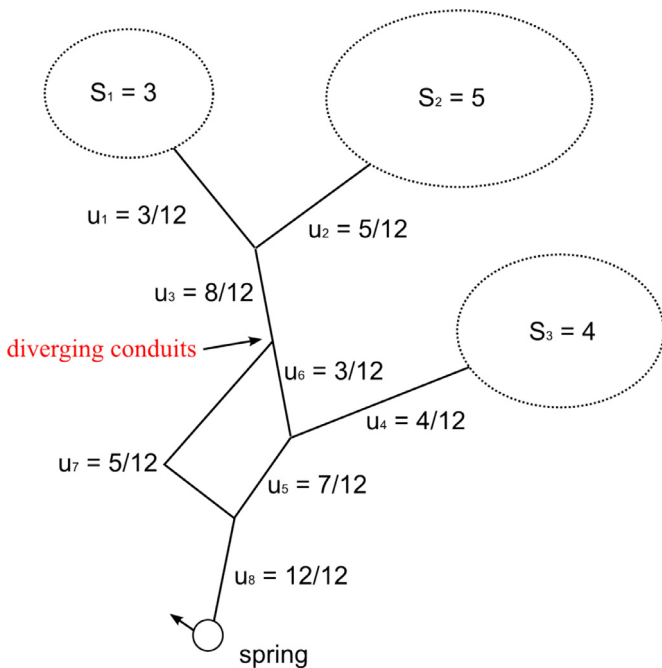


Fig. 10. The order of each conduit u is the ratio between the total catchment surface S that it drains and the total karstic catchment (Eq. 14).

In this section, we consider only one single 3D realization of a synthetic karstic aquifer generated using SKS. It covers an area of 3 km^2 . The 3D mesh is rather coarse to reduce computing time; it is composed of 40 elements in x direction, 30 in y direction and 3 in z direction. The cells are cubic, with a width of 50 m. The conduits are meshed with 1D pipe elements⁴, located along the edges of matrix elements that are meshed as 3D cubes (Fig. 1). The reference parameters are $\alpha = 0.2$, $\beta = 1.5$ (Eq. 13) and $K_M = 10^{-5.5} \text{ [m/s]}$ (the hydraulic conductivity of the matrix). The geometry and distribution of the conduit radius for the reference is displayed in Fig. 11.

The flow is then simulated in transient state using GROUNDWATER [12]. The parameters chosen for the reference allow the representation of a realistic karst behavior with rapid response to pulse event in the recharge function. The synthetic data used for the inverse problem include the head variations in five observation wells placed at the base of the model at coordinates given in Table 1 and the spring discharge at location (1000;500) where the head is prescribed at a value of one meter (head is equal to altitude).

The objective function ϕ is defined as the sum of the squared weighted residuals in the same manner as implemented in PEST [17]:

$$\phi = \sum_{i=1}^{n_o} (w_i \cdot r_i)^2 \quad (15)$$

⁴ Appendix C details the equations used for 1D pipes.

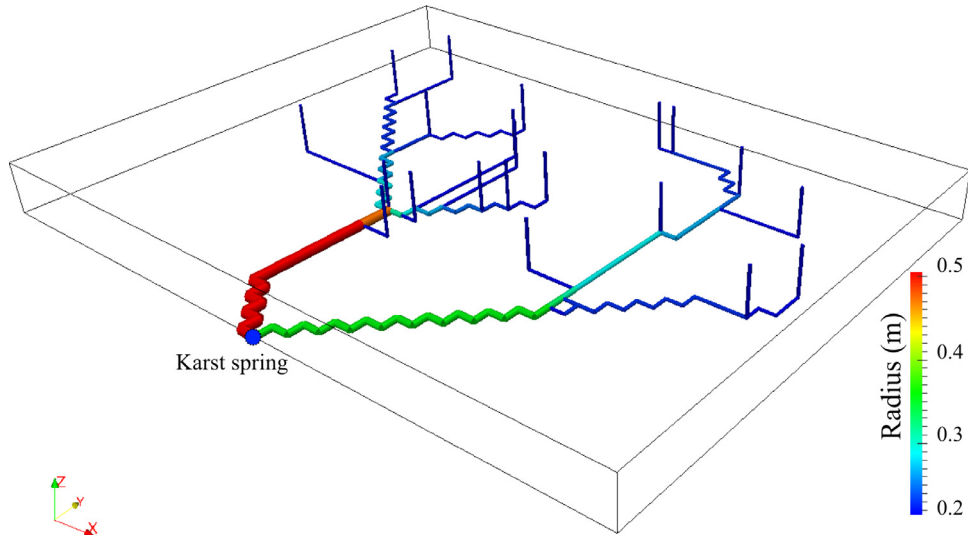


Fig. 11. Distribution of the radius of the conduits in a synthetic case. The radius increases with the order of the conduits. For a visual purpose, the apparent conduit radius has been exaggerated.

Table 1
Observation wells coordinates.

ID	X	Y
H1	1000	750
H2	650	500
H3	650	1000
H4	1350	500
H5	1350	1000

Table 2
Ranges of parameter values for the systematic search. *min/max*: bounds for the parameter; *nb*: number of intervals. α and K_M increments have been computed in a log10 scale.

Parameter	Reference	Min	Max	nb
α [-]	0.2	0.05	1.5	16
β [-]	1.5	0.1	5	14
K_M [m/s]	$10^{-5.5}$	10^{-7}	10^{-4}	16

where n_o is the number of observations, r_i is the residuals of the i th observation and w_i its weight. The observations are taken from 6 different time series. The first one is the spring discharge, and the other 5 are the heads in the observation wells. To give the same weight to the discharge observation and to the head observation, a weight of $1/\sigma$ is given to the discharge observation (where σ is the variance of each time series) and a weight of $1/\sigma \times 1/5$ is given to the head observations because there are 5 observation points. The observation data are presented in Fig. 12: the spring discharge and the hydraulic head observed in boreholes react rapidly with recharge impulses as expected in a realistic karst system.

4.2. Systematic search

In a first step, the objective function ϕ is systematically evaluated for all possible parameter values (α , β , K_M), in order to know precisely whether the problem has a unique or an ensemble of possible solutions and to get some information about the shape of the objective function. The tested parameter ranges are given in Table 2. Note that K_M is sampled in a logarithmic scale, because it can vary over several orders of magnitude. For this step, a total of 3584 flow simulations have been made.

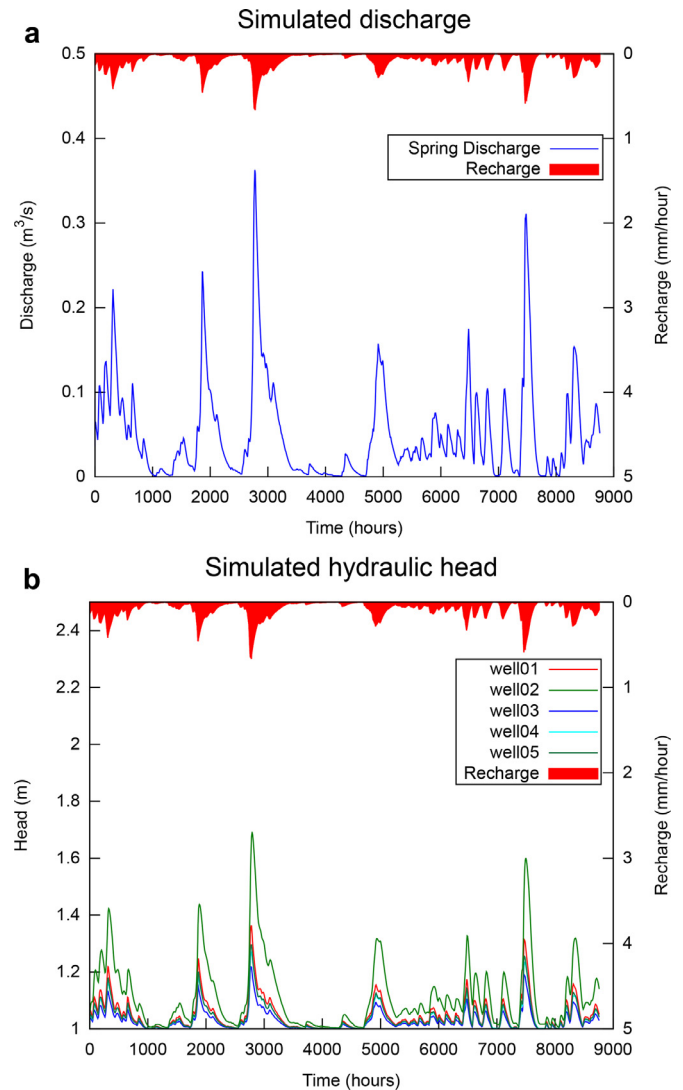


Fig. 12. Reference simulation: (a) simulated spring discharge (b) simulated hydraulic head in observation wells.

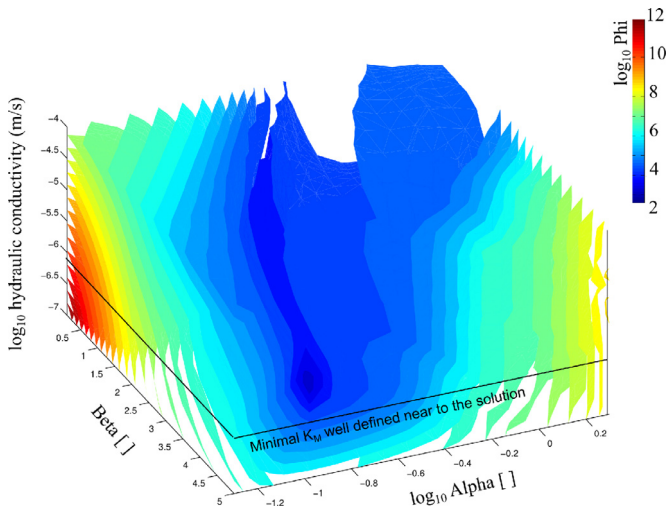


Fig. 13. Systematic search result: isosurfaces of the objective function ϕ as a function of the 3 parameters α , β , and K_m . The axes of the parameter α and K_m are in log scale for improved clarity.

The results are shown in Fig. 13. The objective function ϕ is very sensitive to variations of the hydraulic conductivity of the matrix. Looking in detail at the shape of ϕ in the K_M dimension, one can notice that the value of K_M corresponding to the minimum of ϕ is well defined, but it is far less defined for its high values. This may be explained as follows: on the one hand, if the hydraulic conductivity of the matrix is too low, it leads to too high and unrealistic heads and ϕ is extremely high. On the other hand, if the hydraulic conductivity of the matrix allows the drainage of the diffuse recharge that is infiltrated directly on it, this parameter becomes far less sensitive to the variations.

Another feature of ϕ is that several combinations of α and β can give similar results. The shape of the minimum values of ϕ is quite well defined with respect to K_M , but much less clearly for α and β . This is not surprising because both of these parameters influence the radius r of the conduits. α influences the base radius principally, while β influences the size distribution, i.e. the differences between the biggest and smallest conduits. Therefore, the objective function ϕ has the shape of a valley. Several combinations of α and β values provide a good fit. When α is small the good fits are obtained with large values of β : the smallest conduits are small, and the biggest conduits must be sufficiently large to drain all the system. When α is large, the contrast of size between the smallest and biggest conduits is less strong, i.e. all the conduits have similar sizes, but they are all bigger than in the first case, and the system can be drained equally efficiently.

4.3. Parameter estimation

PEST [17] is one of the most advanced parameter estimation software available for groundwater studies. It is free and open-source. In the present paper it is used as a gradient based optimization technique. The idea is to test the ability of this kind of methods to retrieve the optimal physical parameters for the karst realizations. The way PEST works is conceptually simple:

1. start with an initial guess of the model parameter values
2. run the flow model
3. compute the value of ϕ
4. compute the gradient of ϕ for each parameter (requires several additional runs of the flow model)
5. update the parameters using the Levenberg–Marquardt method [43,44]

6. go to step 2 and repeat until a convergence criterion is reached

To test the ability of gradient based techniques to identify the physical parameters for one karst realization, eight optimization runs have been done, starting approximatively from all the corners of the cube (in the parameter space) defined by the parameters of Table 2. To show the results of these optimization runs, the paths defined by the parameter variations (in the parameter space) have been displayed in Fig. 14a. The value of the objective function ϕ for each parameter combination is displayed as colored balls along the paths. Moreover, the same paths are plotted together with the results of Section 4.2 to show the paths within the plot of ϕ in Fig. 14b.

The results of this test indicate that from the eight optimization runs, one run did not converge toward the right parameters at all, and 5 of them approached very closely the exact reference parameters. The other two were stuck in a local minimum of ϕ , as it can be noted in Fig. 14b. This is a consequence of the shape of ϕ described in the previous section (Section 4.2). The model run time for the forward problem was approximately 5 min, depending on the parameters. The PEST runs needed from 34 to 274 model runs to reach convergence in this case (except one optimization, which failed). Compared to the 3584 simulations of the systematic search, it represents a gain in computational efficiency on the order of 10 to 100.

These results show that gradient search can be an efficient way to estimate the physical parameters of the conduits when the geometry is fixed. As the results of the optimization depend on the starting positions (in the parameter space), one possibility to increase the reliability of this method could be to run several parameter estimations from different starting points and look for the ones with the lowest values of ϕ .

5. Discussion

The results presented here are part of wide research field aiming at better understanding karst aquifers, and providing forecasting and modeling tools for these complex hydrogeological systems. In this work, many assumptions were made to ease the comprehension of the inverse problem applied to karst aquifers. In the following, we summarize and discuss the implications of our results (Section 5.1) and cover also some more general questions about the modeling of karstic systems (Section 5.2).

5.1. Results of this study

The tests that have been performed in the scope of this research show two encouraging results. The first test shows that it is possible to infer some general information such as the number of main conduits, or the conduit radius, using rejection sampling and systematic search methods. The problem associated with this approach is that it is extremely demanding in terms of computer resources. Using a desktop PC (2Gb Ram, 2.8 GHz CPU), it takes about 30 min to run one flow and transport simulation, while the generation of the karst network takes about 5 s. Running the 36,000 simulations required for the first experiment was possible only because we used the high performance linux computing cluster of the University of Neuchtel.

However, the flow model in this first test still remains very simple: it uses an equivalent porous media approach within the conduit elements instead of a more accurate but non-linear flow equations. High Reynolds number values (reaching 3000 in some elements and in some simulations) indicate that Darcy's law is not applicable for those elements and for some simulations. Further research should consider the inertial effects in those elements, but

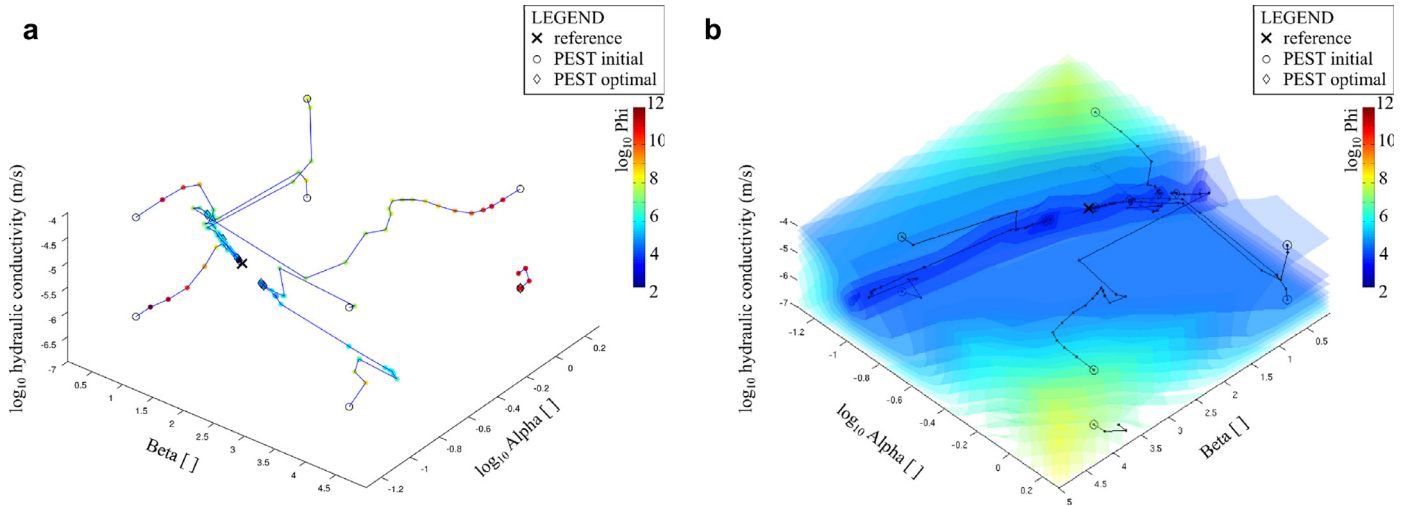


Fig. 14. (a) Eight parameter estimation runs using PEST [17], displayed in the parameter space. 7 runs have given reasonable results and found a parameter set close to the reference. (b) combination of Fig. 13 (with transparency) and (a). We notice that some optimal parameter sets are trapped in local minima of the objective function.

this was not possible in the framework of this project. Nonetheless, the statistical analysis of the error function shows that using tracer test data deeply enhances the ability to infer the geometry and equivalent hydraulic radius of the karst system. Solute transport depends on particle trajectories; head distribution and spring hydrographs are less sensitive to particular flow paths, resulting in a higher degree of uncertainty in the posterior distribution.

This suggests that any information obtained from tracer tests (also natural tracers) should be used to solve the inverse problem in karstic aquifers. The objective function should be a composite function with a strong weight on transport. We observed that with our model setup there was a much larger uncertainty associated with the estimation of the hydraulic conductivity of the limestone matrix in comparison to the conduit radius. We noticed that karst systems with smaller number of drains (2 conduits) were easier to identify than others (e.g. 7 or 14 conduits) when only two tracer tests are available.

The second test showed that it is possible to use standard parameter estimation techniques to optimize the physical properties associated with a geometrical SKS realization. This enhances the computational efficiency of the search by one or two orders of magnitude (10x to 100x less model runs are needed). This second test did not include tracer test simulation, but included more sophisticated physics for the flow problem. Based on the results of the first test, it can be reasonably guessed that the results would benefit of the inclusion of tracer test results to compute the error function. Unfortunately, the direct problem run time would also increase significantly⁵, this is why it was decided not to compute solute transport in this test.

Based on the previous results, we foresee that one possible pragmatic approach to solve the inverse problem in a karstic system could be based on the combination of rejection sampling and gradient based optimization. The procedure could follow three steps: (1) several SKS realizations are generated, (2) the parameters of each one of them are optimized using a gradient based method, and (3) only the ones that provide an acceptable error are kept for uncertainty analysis.

Such a method would have the advantage of reducing the computing time very significantly as compared to a pure rejection sampling, it would however provide only some models that reproduce

the observation, but not a proper set of models within the posterior uncertainty distribution since combining the sampling of the geometry and the optimization step may introduce a statistical bias. Therefore more research is still needed to clarify those problems.

5.2. Open questions and further outlooks

5.2.1. Questions related to the direct problem

In this study, the conduits are simulated as straight pipes, and their radius follows a power law that is dependent of the topological order of the conduits (Eq. 13). It is clear that this approach corresponds to a strong simplification of natural conduits. In reality, the shape of the conduits vary significantly. As a result of anisotropic preferential dissolution, some parts of the same conduit usually present local variation of their radius. Assuming the conduits as cylindrical pipes leads to a strong simplification. These simplifications are necessary because currently no rule exists to characterize these local variations. The use of saturated conduits only results in another significant simplification. Some numerical schemes which allow for the use of variably saturated conduits [16] exist and could be used.

An additional question is related to the importance of simulating the limestone matrix. It would be interesting to investigate in which cases it is possible to exclude the matrix from the problem without losing significant information. This would lead to a simplification of the whole problem, but could drastically enhance the computing efficiency. Using this approach, it could be interesting to test the application of hydrological software for urban conduits network to solve this problem [47]. In the context of inverse problem, it may be possible to use only the conduits for flow during the first estimation of the parameters, and then perform the full coupled simulation only on a few representative models. This could be used as a proxy (approximated physics solver) for flow simulation.

Saller et al. [52] use MODFLOW-CFP and transfer functions to simulate the interactions between the matrix and conduits. They identify the difficulty in correctly assessing the values of these transfer functions, but show also that this gives a realistic behavior to these interactions. In our study, we use a direct hydraulic connection between matrix and conduits. Further studies should be achieved to include transfer functions in the inverse problem framework. This would probably results in an even more complex inverse problem because more parameters would need to be identified. But, on the other hand, the use of transfer

⁵ Appendix A details a model-run workflow to minimize the necessary cpu time of direct problems.

functions could enhance the numerical efficiency because the contrasts of hydraulic conductivity would be smoother, and the direct flow problem would probably reach convergence more rapidly, reducing the CPU time needed.

5.2.2. Questions related to the inverse problem

Several aspects of the inverse problem have not been taken into account in the present work. For example, the recharge model used in this study is based on a reservoir model. The identification of its parameters was not included in the inverse problem framework. As Weber et al. [61] show, the recharge model has a very strong impact on spring hydrographs for real applications. Similarly, the matrix and the radius of the conduits have been described with a very small number of parameters, while we expect that the matrix will be heterogeneous with parameters varying in space and we expect the distribution of the conduit diameters to be much more complex than the simple power-law proposed here. All these additional complexity will need to be faced when dealing with a real case.

In terms of methodology, a method that could be used to increase the numerical efficiency of the whole inverse framework is to employ approximate physics solvers (so called “proxies”) and classification techniques as a preliminary step to select candidate models (parameter sets) and to minimize the number of runs of the highly demanding and accurate direct problem [55]. A further technique is proposed by Ginsbourger et al. [24] who used kriging to interpolate the error function using proxies. Such an approach could be used for karst aquifers as well, but additional research is required to identify good proxies for karst aquifers.

Conclusion

In this paper, several aspects of the inverse problem in karst modeling have been studied. Our numerical experiments did not aim at solving the inverse problem, but rather at providing a better understanding of the underlying challenges. Indeed, the estimation of the 3D geometry of a karstic network and of the corresponding physical parameters can be impossible to achieve by a manual trial and error approach. Therefore, these models should benefit from available automatic parameter estimation techniques. Furthermore, the inverse problem needs to be applied with the aim to evaluate the uncertainty related to these models and their parameters.

The numerical experiments described in this paper show two very encouraging results.

The first test has shown that both physical and geometrical parameter can be identified pretty accurately in an inverse framework if adequate data is available. By repeating the experiment a very large number of times with different references having different number of conduits, different matrix permeability and different conduit radius, we could statistically evaluate how efficiently the inverse method can retrieve the features of the reference from the head or tracer data. This test indicated that data from tracer tests had a very strong impact on the identification of the geometrical properties of the karst network.

The second test has shown the possibility of improving the numerical efficiency of the inverse algorithm by accelerating the search for optimal parameters when the geometry is fixed using gradient-based methods. The model that has been used for this test considers more complicated physics than the one of the first test. A systematic analysis of the objective function ϕ shows that several combinations of the parameters that influence the radius of the conduits (α and β) can give more or less equally satisfactory results. The hydraulic conductivity of the matrix is very sensitive to the low conductivity values, but as soon as it is sufficient to let the recharge infiltrate and let the water reach the conduits, ϕ is far

less sensitive to its variations. This test also shows that PEST [17] can be used efficiently to solve this problem.

These results indicate that one could solve the inverse problem in karstic aquifers in a pragmatic manner using a two steps approach. First, a set of many different karst geometrical realizations are generated using the pseudo-genetic methodology [8], and secondly for each one of them the best physical parameters are found using a parameter estimation technique. The resulting set of simulations can be used for further analysis.

In both tests, we used a combination of already existing tools to simulate flow and transport in karst aquifers with distributed parameter models. The karstic conduits are meshed as 1D pipes that follow the edges of the 3D elements used for the matrix. Turbulent flow equations are used to simulate flow in the conduits. The radius of the conduits depends on their hierarchical order. Major conduits have largest radiuses, while secondary conduits are smaller. The conduits are hierarchically classified by the ratio of their catchment surface over the total catchment surface. The recharge function applied on the inlets also depends on their catchment surface. The greatest part of recharge is applied directly to the inlets and only a small percentage is applied on the rest of the model. This allows for an approximation of the effect of epikarst in conveying water to the inlets. We argue that this methodology is pretty general and could be used for many karstic systems.

Finally, if we consider again the question stated in the title of the present study “is it possible to identify karst conduit networks geometry and properties from hydraulic and tracer test data?”, the answer that is suggested is “Yes, it seems possible to bracket the values of the radius of the conduits within a reasonable range, provided that enough computing power and sufficient data are available”. In practice, this study shows very promising results, but it is based on a limited set of numerical experiments and restrictive assumptions. Further research is needed to enhance the inverse framework, and especially the solving of the direct problem. As stated in the discussion, maybe the use of approximate physics to simulate the direct problem within the inverse problem framework could allow to highly speed up the search of valid karst realizations, and therefore allow such a methodology to be applied on real size aquifer problems.

Acknowledgments

The funding of this research was provided by the Swiss National Foundation for Research, on fund no. P2NEP2-151935. Thanks also to *E-Dric.ch* company for providing their watershed management software *RS2012* for this research. Partial funding of this research was provided also by Schlumberger Water Services.

Appendix A. Model run workflow

To be used in an inverse problem framework, the model run workflow is defined in order to minimize the total necessary CPU time. The workflow must take into account transient flow conditions, and transient transport. When studying karst aquifers, several tracer tests may be performed. Therefore, the model should match all of them. Unfortunately, the transport simulation, especially with strong heterogeneity in the flow medium, may lead to very high computational times. The model run workflow is intended to reduce the necessary CPU time as much as possible, when simulating flow and transport at transient state. It is separated in three main steps:

1. steady state flow simulation
2. long transient state (years) flow simulation
3. short transient state (months) flow and transport simulations for each tracer

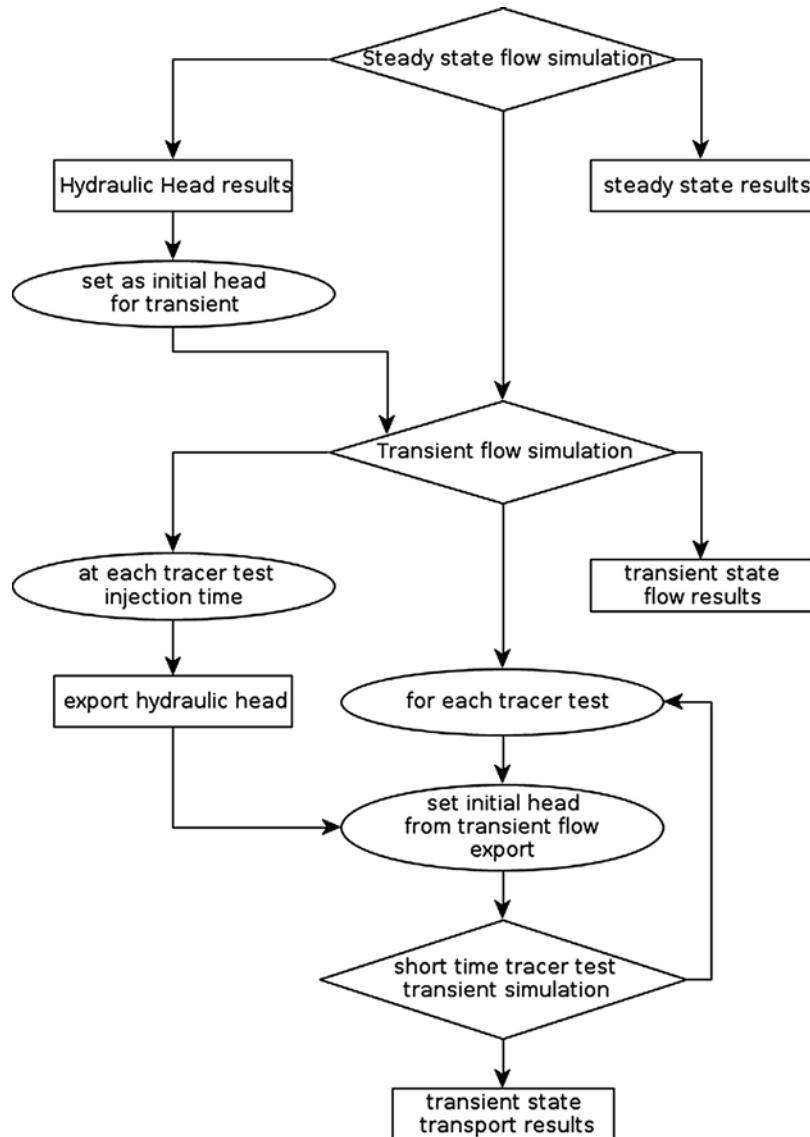


Fig. A.15. Flowchart of the model run.

The first step of the model run is a steady state flow simulation using the average flow conditions (i.e. the average discharge of the springs) as input for the model. The steady state hydraulic head result is then used as initial conditions for the transient flow simulation. The steady state simulation has to be run in the same hydrological conditions as the beginning of the transient simulation, e.g. if the transient simulation starts in a low-flow period, the steady state simulation must be run under low flow condition too, and vice-versa.

The transient simulation is done for the whole period of interest, which must include the periods of the tracer tests, and which may last a few years. During this simulation, the hydraulic head results at every tracer test injection time is stored to initialize the flow field of the short flow and transport simulations. The transport simulations are performed only over a short period of time (a few days or weeks, depending on the duration of the experiments), because they are heavier to compute. The advantage of this workflow is that the tracer simulations are consistent with the whole transient flow simulation, without needing to solve the whole transient simulation (years) including transport. Fig. A.15 shows the flowchart of the model run.

Appendix B. Simulating flow and transport by equivalent medium “karst” elements

The first tests with the hierarchical conduit networks have been done using an equivalent porous medium (EPM) approach, even if this does not allow the simulation of the full complex karstic flow (e.g. turbulent flow in the conduits). Still, some authors have obtained satisfactory results in some cases such as Scanlon et al. [54] who use equivalent properties for entire regions of their model. In this section, the EPM approach is used by assigning specific flow and transport properties to the karst conduit elements, similarly to Worthington [62]. This method has the advantage of being simple and fast to solve as compared to the use of discrete elements with non linear flow equations (Section Appendix C).

The properties of the conduit elements are derived from the radius of the conduits and the mesh size in the following manner: the equivalent porosity ϕ_{eq} (-) is computed by estimating the volume of void in the cell divided by the volume of the cell:

$$\phi_{eq} = \frac{V_{void}}{V_{cell}} = \frac{V_{conduit} + (V_{cell} - V_{conduit}) \cdot \phi_{matrix}}{V_{cell}} \quad (B.1)$$

where $V_{conduit}$ is the volume of the conduit (m^3), i.e. $\pi r^2 d$, with d the length in the cell (m), ϕ_{matrix} is the porosity of the matrix (-).

The equivalent permeability is computed such that the total discharge Q_{tot} (m^3/s) flowing through one cell in a given direction is identical in the equivalent porous medium or in the pipe plus matrix model. The total discharge is given by:

$$Q_{tot} = Q_m + Q_c \quad | \quad Q_c = -\frac{\pi r^4}{8} \frac{\rho g}{\mu} \nabla H,$$

$$Q_m = -\kappa_m \frac{\rho g}{\mu} \cdot \frac{(A - \pi r^2)}{A} \nabla H \quad (B.2)$$

where $A[m^2]$ is the area of the cell side crossed by the flux, πr^2 the area of the conduit, $Q_c[m^3/s]$ is the discharge through the conduit that crosses the element (using the Poiseuille law), $Q_m[m^3/s]$ is the discharge flowing through the matrix surrounding the conduit, κ_m is the permeability of the matrix (m^2), g is the acceleration of the gravity ($m \cdot s^{-2}$), ρ is the water density ($kg \cdot m^{-3}$) and μ is the water dynamic viscosity ($kg \cdot m^{-1} \cdot s^{-1}$) and ∇H is the hydraulic gradient. Therefore the total flux q_{tot} (m/s) is equal to:

$$q_{tot} = \frac{Q_{tot}}{A} = -\left[\frac{\pi r^4}{8A} + \kappa_m \frac{(A - \pi r^2)}{A} \right] \frac{\rho g}{\mu} \nabla H = -\kappa_{eq} \frac{\rho g}{\mu} \nabla H \quad (B.3)$$

The equivalent permeability of the cell κ_{eq} (m^2) is then given by:

$$\kappa_{eq} = \kappa_m + \frac{\pi r^2}{A} \cdot \frac{r^2}{8} - \kappa_m \quad (B.4)$$

The use of an equivalent porous medium approach is a strong approximation, especially for transport problems, as we need to put unrealistically low porosity to the karst elements (the elements containing a karst conduit) so that the pore velocity becomes satisfactory in terms of transport [37].

Appendix C. Simulating flow and transport in 1D pipes embedded in a 3D porous matrix

The whole model is meshed using a regular grid (Section 2.2). Pipes elements are located on the edges of the 3D elements (hexahedrons) of the karstic matrix. There is a direct hydraulic connection between the pipes and the matrix, as the nodes of both types of elements are the same.

To model the flow in the conduits two different approaches can be used: laminar and turbulent flow. The linear flow equation used in pipes is the Darcy-Poiseuille formula, the hydraulic conductivity for the pipes elements in this case is:

$$\kappa = \frac{\rho g r^2}{\mu 8} \quad (C.1)$$

The flow equation used in pipes for turbulent flow is the Manning-Strickler (Turbulent flow conditions). The hydraulic conductivity for the pipes is therefore:

$$\kappa = \frac{\phi^{\frac{r}{2}}}{\|\nabla H\|} \quad (C.2)$$

where ϕ is the friction coefficient ($m^{1/3}s^{-1}$). Both cases are used with the same law for flow ($\vec{v} = -K\nabla H$) where only the definition of K varies from one case to the other ($K = \kappa \frac{\rho g}{\mu}$).

Appendix D. Transport simulation

The injection of the tracer is modeled as an initial concentration at the beginning of the transport simulation. The initial flow field

of the simulation is extracted from the transient long period flow simulation. In this way, all the transport simulations are consistent with the long period flow simulation. Ideally, the transient flow field of the transient flow simulation should be used as flow field for the transport simulation, to reduce the computational time, but this is not possible, because the time-steps evolution depends also from the transport simulation.

According to many authors [3–5], it is necessary to pay attention to two parameters: the Peclet number Pe and the Courant number Cr [13]. The Peclet number is defined as:

$$Pe = \frac{v dx}{D} \quad (D.1)$$

where v is the flow velocity, i.e. the pore velocity v_p multiplied by the porosity ϕ ($v = v_p \phi$). dx is the mesh dimension in the flow direction (in this case $dx = dy = dz$) and D is the dispersion coefficient (m^2/s). The Courant number is defined as:

$$Cr = \frac{v \delta t_{max}}{dx} \quad (D.2)$$

where δt_{max} is maximal time step size (s) for the transient simulation. The numerical solution can be considered stable if these two conditions are satisfied:

$$Pe < 2, \quad Cr < 1 \quad (D.3)$$

These two inequalities can be expressed in order to isolate the dispersion D and the maximum time step δt_{max} :

$$\frac{v dx}{2} < D, \quad \delta t_{max} < \frac{dx}{v} \quad (D.4)$$

The dispersion is a parameter that depends on the scale of the problem, as it can be used to assume the dispersion of a pollutant induced by pore-scale heterogeneities in an equivalent porous medium. As explained in Rausch et al. [49] the effect of other kinds of heterogeneities (like karst features in this case) may strongly influence the behavior of the solute in the simulation. In this case, the dispersion has to be interpreted as the dispersion occurring at cell scale, and depends therefore strongly on the mesh dimensions.

Moreover, following the conditions of Eq. D.4, it appears that particular attention has to be paid to mesh size and on the values to the dispersion coefficient. The time discretization is also very important. However, when simulating karst aquifers, with a very strong heterogeneity, it can be difficult to compute the value of the maximal flow velocity in the medium a priori, which can be very fast in some cases. To guess these parameters before running the full transient transport simulation, a steady state flow simulation can be performed in high flow conditions.

References

- [1] Annable WK. Numerical analysis of conduit evolution in karstic aquifers Ph.D. thesis. University of Waterloo, Canada; 2003.
- [2] Bakalowicz M. Karst groundwater: a challenge for new resources. *Hydrogeol J* 2005;13(1):148–60.
- [3] Bear J. Dynamics of Fluids in Porous Media. Dover publications; 1972.
- [4] Bear J, Jacobs M. On the movement of water bodies injected into aquifers. *J Hydrol* 1965;3:37–57.
- [5] Bear J, Tsang C-F, de Marsily G. Flow and contaminant transport in fractured rock. Academic Press, San Diego, CA (United States); 1993.
- [6] Borghi A. Modélisation 3d d'un aquifère karstique, la Vallée des Ponts-de-Martel et la source de la Noiraigue Master's thesis. University of Neuchâtel, Switzerland; 2008.
- [7] Borghi A, Renard P, Fournier L, Negro F. Stochastic fracture generation accounting for the stratification orientation in a folded environment based on an implicit geological model. *Eng Geol*, 187; 2015. p. 135–42.
- [8] Borghi A, Renard P, Jenni S. A pseudo-genetic stochastic model to generate karstic networks. *J Hydrol* 2012;414–415(0):516–29.
- [9] Caers J, Hoffman T. The probability perturbation method: A new look at bayesian inverse modeling. *Math Geol* 2006;38:81–100.
- [10] Carrera J, Neuman SP. Estimation of aquifer parameters under transient and steady-state conditions: 1. maximum likelihood method incorporating prior information. *Water Resour Res* 1986;22:199–210.

- [11] Collon-Drouaillet P, Henrion V, Pellerin J. An algorithm for 3d simulation of branchwork karst networks using horton parameters and \rightarrow application to a synthetic case. *Geol Soc London Spec Publ*, 370; 2012. p. 295–306. URL <http://sp.lyellcollection.org/content/early/2012/06/21/SP370.3.abstract>.
- [12] Cornaton F. Ground water: A 3-d ground water and surface water flow, mass transport and heat transfer finite element simulator. Internal report, University of Neuchâtel; 2007.
- [13] Courant R, Friedrichs K, Lewy H. On the partial difference equations of mathematical physics. *IBM J Res Dev* 1967;11(2):215–34.
- [14] De Marsily G. Quelques réflexions sur l'utilisation des modèles en hydrologie. *Rev des sci l'eau / J Water Sci* 1994;7:219–34.
- [15] De Rooij R. Towards improved numerical modeling of karst aquifers: coupling turbulent conduit flow and laminar matrix flow under variably saturated conditions Ph.D. thesis. University of Neuchâtel, Switzerland; 2007.
- [16] De Rooij R, Perrochet P, Graham W. From rainfall to spring discharge: Coupling conduit flow, subsurface matrix flow and surface flow in karst systems with a discrete-continuum model Paper in preparation 2013.
- [17] Doherty J, Brebber L, Whyte P. Pest: Model-independent parameter estimation, 122. Corinda, Australia: Watermark Computing; 1994.
- [18] Dreybrodt W, Gabrovsek F, Douchko R. Processes of speleogenesis: a modelling approach, vol. 4. Zalozhba; 2005.
- [19] Filipponi M, Jeannin PY, Tacher L. Evidence of inception horizons in karst conduit networks. *Geomorphology* 2009;106:86–99. Sp. Iss. SI; URL <Go to ISI>://000264975700009.
- [20] Ford D, Williams P. Karst hydrogeology and geomorphology. Chichester, England: John Wiley and Sons, Ltd.; 2007.
- [21] Fournillon F, Viseur S, Arfib B, Borgomano J. Insights of 3d geological modelling in distributed hydrogeological models of karstic carbonate aquifers. *Adv Res Karst Media* 2010:257–62.
- [22] Gabrovsek F, Dreybrodt W. Karstification in unconfined limestone aquifers by mixing of phreatic water with surface water from a local input: a model. *J Hydrol* 2010;386:130–41. URL <Go to ISI>://000278577000011.
- [23] Ghasemizadeh R, Hellweger F, Butscher C, Padilla I, Vesper D, Field M, et al. Review: groundwater flow and transport modeling of karst aquifers, with particular reference to the north coast limestone aquifer system of puerto rico. *Hydrogeol J* 2012;1–21. <http://dx.doi.org/10.1007/s10040-012-0897-4>.
- [24] Ginsbourger D, Rossopoff B, Pirot G, Durrande N, Renard P. Distance-based kriging relying on proxy simulations for inverse conditioning. *Adv Water Resour* 2013;52(0):275–91. URL <http://www.sciencedirect.com/science/article/pii/S0309170812003016>.
- [25] Goldscheider N, Drew D. Methods in Karst Hydrogeology: IAH: International Contributions to Hydrogeology, 26. Taylor & Francis; 2007.
- [26] Groves C, Howard A. Early development of karst systems: 1. preferential flow path enlargement under laminar flow. *Water Resour Res* 1994;30(10):2837–46.
- [27] Hadamard J. Le problème de Cauchy et les équations aux dérivées partielles linéaires hyperboliques; 1932.
- [28] Henrion V, Pellerin J, Caumon G. A stochastic methodology for 3D cave system modeling. Proceedings of eighth Geostatistical Geostatistics Congress, Vol 2; 2008. p. 525–34.
- [29] Hu C, Hao Y, Yeh T-C J, Pang B, Wu Z. Simulation of spring flows from a karst aquifer with an artificial neural network. *Hydrol Process* 2008;22(5):596–604. <http://dx.doi.org/10.1002/hyp.6625>.
- [30] Hu LY. Gradual deformation and iterative calibration of gaussian-related stochastic models. *Math Geol* 2000;32:87–108.
- [31] Jaquet O, Siegel P, Klubertanz G, Benabderhamane H. Stochastic discrete model of karstic networks. *Adv Water Resour* 2004;27.
- [32] Jeannin DP-Y. Modeling flow in phreatic and epiphreatic karst conduits in the hoelloch cave(muotatal, switzerland). *Water Resour Res*, 37, no 2; 2001. p. 191–200.
- [33] Király L. Large scale 3-D groundwater flow modelling in highly heterogeneous geologic medium; 1988. p. 761–75.
- [34] Király L. Karstification and groundwater. *Speleogenesis and Evolution of Karst Aquifers*; 2003. p. 155–90. see http://link.springer.com/chapter/10.1007%2F978-94-009-2889-3_38
- [35] Király L, Burger A, Dubertret L. État actuel des connaissances dans le domaine des caractères physiques des roches karstiques [Report on the present knowledge of the physical characters of karstic rocks], vol. Series 8 - Number 3; 1975. p. 53–67.
- [36] Király L, Perrochet P, Rossier Y. Effect of the epikarst on the hydrograph of karst springs: a numerical approach. *Bull d'hydrogéol* 1995;14.
- [37] Kresic N, Stevanovic Z. Groundwater Hydrology of Springs ,Engineering, Theory, Management and Sustainability. Springer; 2009.
- [38] Kuniansky E, Shoemaker W. Usgs releases conduit flow process (cfp) for modflow-2005. 2008 Joint Meeting of The Geological Society of America, Soil Science Society of America, American Society of Agronomy, Crop Science Society of America, Gulf Coast Association of Geological Societies with the Gulf Coast Section of SEPM 2008.
- [39] Novakowski KS, Lapcevic PA, Sudicky EA. Groundwater flow and solute transport in fractured media. Delleur JW, editor. *The Handbook of Groundwater Engineering*, Second edition. CRC Press; 2006. p. 1–43. <http://dx.doi.org/10.1201/9781420006001.ch20>.
- [40] Larocque M, Banton O, Ackerer P, Razack M. Determining karst transmissivities with inverse modeling and an equivalent porous media. *Ground Water* 1999;37(6):897–903. <http://dx.doi.org/10.1111/j.1745-6584.1999.tb01189.x>.
- [41] Larocque M, Banton O, Razack M. Transient-state history matching of a karst aquifer ground water flow model. *Ground Water* 2000;38(6):939–46. <http://dx.doi.org/10.1111/j.1745-6584.2000.tb00694.x>.
- [42] Mangin A. Contribution à l'étude hydrodynamique des aquifères karstiques Ph.D. thesis. Université de Dijon; 1975.
- [43] Marquardt D. An algorithm for least-squares estimation of nonlinear parameters. *J Soc Ind Appl Math* 1963;11(2):431–41.
- [44] More J. The levenberg-marquardt algorithm: implementation and theory. *Numerical Analysis* 1978:105–16.
- [45] Ofterdinger U, Renard P, Loew S. Hydraulic subsurface measurements and hydrodynamic modelling as indicators for groundwater flow systems in the Rotondo granite, central alps (Switzerland). *Hydrol Process* 2013. see <http://onlinelibrary.wiley.com/doi/10.1002/hyp.9568/abstract;jsessionid=0AD49002EF4020E7D429892AE4496A09.f02t02>
- [46] Panagopoulos G. Application of modflow for simulating groundwater flow in the trifilia karst aquifer, greece. *Environ Earth Sci* 2012;67:1877–89. <http://dx.doi.org/10.1007/s12665-012-1630-2>.
- [47] Peterson EW, Wicks CM. Assessing the importance of conduit geometry and physical parameters in karst systems using the storm water management model (swmm). *J Hydrol* 2006;329:294–305. Times Cited: 2; URL <Go to ISI>://WOS:000240888300023.
- [48] RamaRao BS, Lavenue M, de Marsily G, Marietta MG. Pilot point methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields: 1. theory and computational experiments. *Water Resour Res* 1995;31:475–93.
- [49] Rausch R, Schäfer W, Therrien R, Wagner C, et al. Solute transport modelling: an introduction to models and solution strategies. Gebrüder Borntraeger Verlagsbuchhandlung; 2005.
- [50] Ronayne MJ, Gorelick SM. Effective permeability of porous media containing branching channel networks. *Phys Rev E Stat Nonlinear Soft Matter Phys* 2006;73:026305. Times Cited: 0; URL <Go to ISI>://INSPEC:8831472.
- [51] Sagar B, Yakowitz S, Duckstein L. A direct method for the identification of the parameters of dynamic nonhomogeneous aquifers. *Water Resour Res* 1975;11(4):563–70. <http://dx.doi.org/10.1029/WR011i004p00563>.
- [52] Saller S, Ronayne M, Long A. Comparison of a karst groundwater model with and without discrete conduit flow. *Hydrogeol J* 2013;21(7):1555–66. <http://dx.doi.org/10.1007/s10040-013-1036-6>.
- [53] Sambridge M, Mosegard K. Monte carlo methods in geophysical inverse problems. *Rev Geophys* 2002;40:1009.
- [54] Scanlon BR, Mace RE, Barrett ME, Smith B. Can we simulate regional groundwater flow in a karst system using equivalent porous media models? case study, barton springs edwards aquifer, usa. *J Hydrol* 2003;276(14):137–58. URL <http://www.sciencedirect.com/science/article/pii/S0022169403000647>.
- [55] Scheidt C, Caers J. Uncertainty quantification in reservoir performance using distances and kernel methodsapplication to a west africa deepwater turbidite reservoir. *SPE J* 2009;14(4):680–92.
- [56] Sethian JA. Level Set Methods and Fast Marching Methods Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Materials Science (re-edition); 2008. <http://www.cambridge.org/us/catalogue/catalogue.asp?isbn=0521645573>
- [57] Tarantola A. Inverse Problem Theory and methods for parameter estimation. Philadelphia SIAM; 2005. ISBN 978-0-898715-72-9.
- [58] Tonkin M, Doherty J, Moore C. Efficient nonlinear predictive error variance for highly parameterized models. *Water Resour Res* 2007;43. Times Cited: 0; URL <Go to ISI>://WOS:000248291000005.
- [59] von Neumann J. Monte Carlo methods. National Bureau of Standards. Various techniques used in connection with random digits; 1951. p. 3638. chap. 12.
- [60] Vuilleumier C, Borghi A, Renard P, Ottowitz D, Schiller A, Supper R, et al. A method for the stochastic modeling of karstic systems accounting for geophysical data: an example of application in the region of Tulum, Yucatan Peninsula (Mexico). *Hydrogeol J* 2013;21(3):529–44. <http://dx.doi.org/10.1007/s10040-012-0944-1>.
- [61] Weber E, Jeannin P-Y, Malard A, Vouillamoz J, Jordan F. A pragmatic simulation of karst spring discharge with semi-distributed models. advantages and limits for assessing the effect of climate change. In: Actes du 13e Congrès national de SpActes du 13e Congrès national de Spéléologie; 2012.
- [62] Worthington S. Characterization of the mammoth cave aquifer. In: Significance of caves in watershed management and protection in florida; 2003.
- [63] Worthington S, Ford D, Beddows P, Klimchouk A, Ford D, Palmer A, et al.. Porosity and permeability enhancement in unconfined carbonate aquifers as a result of solution; 2000. p. 527.
- [64] Worthington SRH, Smart CC, Ruland WW. Assessment of groundwater velocities to the municipal wells at Walkerton. In: Proceedings of the 2002 joint annual conference of the canadian geotechnical society and the canadian chapter of the IAH; 2002.