



HAL
open science

KarstNSim: A graph-based method for 3D geologically-driven simulation of karst networks

Augustin Gouy, Pauline Collon, Vincent Bailly-Comte, Eric Galin, Christophe Antoine, Benoît Thebault, Philippe Landrein

► **To cite this version:**

Augustin Gouy, Pauline Collon, Vincent Bailly-Comte, Eric Galin, Christophe Antoine, et al.. KarstNSim: A graph-based method for 3D geologically-driven simulation of karst networks. Journal of Hydrology, 2024, 632, pp.130878. 10.1016/j.jhydrol.2024.130878 . hal-04468018v1

HAL Id: hal-04468018

<https://hal.univ-lorraine.fr/hal-04468018v1>

Submitted on 20 Feb 2024 (v1), last revised 1 Mar 2024 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NoDerivatives 4.0 International License

KarstNSim: a graph-based method for 3D geologically-driven simulation of karst networks

Augustin Gouy¹, Pauline Collon¹, Vincent Bailly-Comte^{2,3}, Eric Galin⁴, Christophe Antoine¹, Benoît Thebault¹, and Philippe Landrein⁵

¹Université de Lorraine, CNRS, GeoRessources, F-54000 Nancy, France

²BRGM, Univ. Montpellier, Montpellier, France

³G-eau, INRAE, CIRAD, IRD, AgroParisTech, Supagro, BRGM, Montpellier, France

⁴UCBL, CNRS, INSA Lyon, LIRIS, UMR5205, France

⁵Andra, Centre de Meuse / Haute-Marne, RD 960, 55290 Bure

Abstract In karst aquifers, groundwater flow is highly influenced by the interconnected underground cavities and conduits that form the karst network. Modeling karst flows requires the use of spatially distributed approaches accounting for these networks. Their exploration is, however, often complex, and mapping them using indirect methods such as geophysical ones has proven challenging. To overcome these limitations, stochastically simulating discrete karst networks should account for the uncertainties on conduit position and geometry. Only a few existing methods can reproduce realistic and diverse karst morphologies. In this work, we propose a new algorithm, KarstNSim, for simulating discrete karst networks, that incorporates field data to generate a range of possible karst network geometries. It relies on the computation of the shortest path between the inlets and outlets of the network with the use of an anisotropic cost function defined on an n-nearest neighbor graph conformal to geological and structural heterogeneities. This cost function represents the physico-chemical processes that govern speleogenesis – such as erosion and chemical weathering – providing simplified control over the geometry of the generated networks. Our approach reproduces the vadose-phreatic partition visible in the karst networks, by generating sub-vertical conduits in the unsaturated zone and sub-horizontal ones in the saturated part. It encompasses geological parameters such as inception surfaces, fractures, permeability, and solubility of layers, along with considering the hydrological context of recharge by assigning relative weights to the inlets. We simulate various synthetic models to demonstrate the influence of different input parameters on the spatial organization of the past and present karst flows. We also apply KarstNSim to a real case study —the Ribeaucourt system (Meuse, France)— and use available data to generate realistic karst networks. The lack of knowledge on the Ribeaucourt network geometry is alleviated by the use of multiple scenarios corresponding to different sets of weights in the cost function, resulting in a variety of network morphologies which represent the uncertainty on the real geometry. The geometrical and topological metrics computed on the simulated networks generally overlap with those of real ones, suggesting the reliability of the method.

Keywords

stochastic simulations
karst network
morphology
cost function
shortest path

INTRODUCTION

The formation of karst networks mostly occurs in carbonate formations where it controls groundwater flow dynamics in the massif. Implicit and explicit approaches of karst conduit network representation exist for modeling groundwater flow in such formations. Implicit methods include global or lumped approaches, such as rainfall discharge models (e.g., Mazzilli et al., 2019) and reservoir models (e.g., Thiéry, 1982). They directly translate recharge data into discharge data. It enables the computation of the impluvium surface and characterization of the water balance, but it does not allow for the computation of hydraulic head fields and groundwater flow fields. Other implicit methods include Equivalent Porous Medium (EPM) and Double Continuum models, which are both distributed, but consider the karst features as homogeneously distributed in the medium (Ghasemizadeh et al., 2012).

In contrast, explicit approaches insert discrete karst networks

in the model, allowing the computation of local variations of the hydraulic heads and the simulation of flow rate fields (e.g., Kiraly, 1988; de Rooij, 2007; Borghi et al., 2016; Sivelle et al., 2020; Duran and Gill, 2021). Yet, those models are sensitive to the number and position of the conduits (Borghi et al., 2016; Duran and Gill, 2021), which are most of the time unknown. Therefore, it is often necessary to generate the karst network before conducting a physically-based karst flow simulation.

Yet, explicit simulation of karst networks remains challenging. Karstification depends on various physico-chemical processes and occurs in different geological and hydrogeological contexts (fracturing, inception horizons (Filipponi et al., 2009), infiltration conditions...) and over variable periods of time (from a few years to millions of years) (Mangin, 1974). It produces a wide variety of karst network patterns, which were studied and schematically classified by Palmer (1991). Those patterns were further specified and associated with different

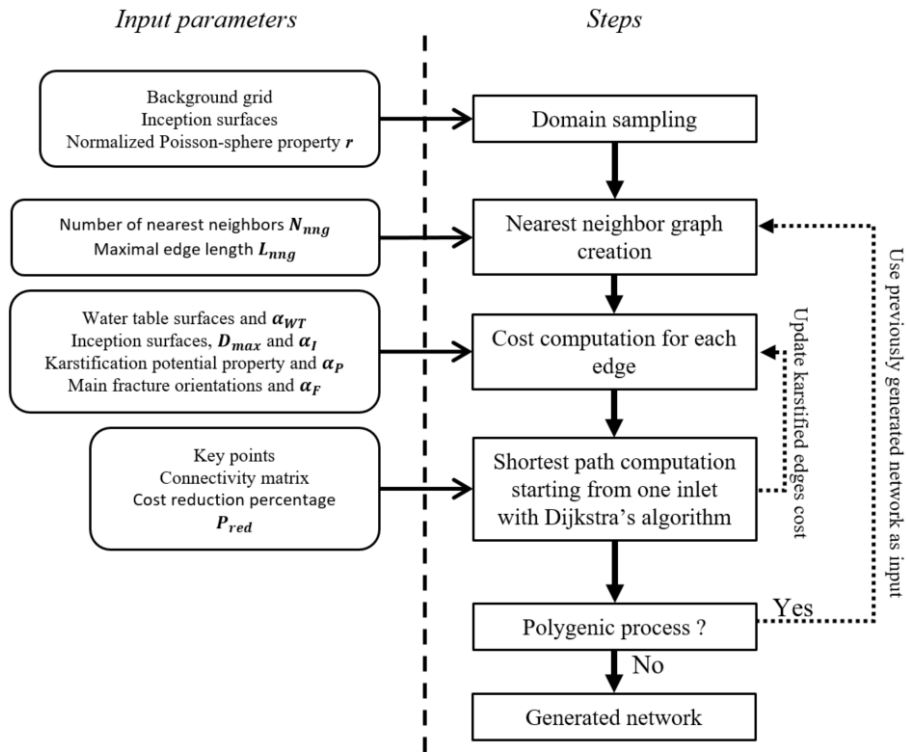


Figure 1 Workflow of the KarstNSim methodology.

hydrogeological contexts by [Jouves et al. \(2017\)](#). As a result, real karst networks do not correspond to a specific pattern, but rather a combination of patterns.

Different types of algorithms have been used to model karst networks in epigenic contexts: process-based methods (e.g., [Siemers and Dreybrodt, 1998](#)), geostatistical methods (e.g., [Fournillon et al., 2010](#)) and pseudo-genetic methods (e.g., [Jaquet et al., 2004](#)), each having their advantages and disadvantages. In general, process-based methods are the only ones capable of replicating the kinetics of speleogenesis, whereas geostatistical and pseudo-genetic methods directly produce the outcome of a karstification phase. However, the latter approaches are less computationally intensive and require fewer input parameters than process-based methods. Overall, pseudo-genetic algorithms offer the most flexibility to conditioning data, which can be of varying nature: direct data such as cave maps, indirect data such as tracer tests, location of inlets or areas of diffuse infiltration, position of springs, and observation of surface fracturation. This flexibility arises from the principles of those methods: pseudo-genetic algorithms use numerical tips to mimic the effects of the laws governing speleogenetic processes but not the processes themselves. Most often, they use a cost function to compute the path of least hydraulic resistance between an inlet and an outlet of the network ([Henrion et al., 2008](#); [Borghi et al., 2012](#); [Collon-Drouaillet et al., 2012](#); [Paris et al., 2021](#)).

Several methods exist to compute karst networks with a pseudo-genetic algorithm. The Stochastic Karst Simulation method (SKS) of [Borghi et al. \(2012\)](#), later reused and modified by [Luo et al. \(2021\)](#) and [Fandel et al. \(2022\)](#), is a grid-based approach which uses a Fast-Marching algorithm ([Sethian, 1996, 2001](#)) to solve the shortest path problem. While the method offers flexible control over the morphology of the networks, the structured grid on which all properties are stored may involve aliasing effects depending on the spatial resolution of the grid, especially in the case of a fractured

medium. The structured grid may also not be conformal to heterogeneities.

[Henrion et al. \(2008\)](#), and later [Collon-Drouaillet et al. \(2012\)](#) proposed to generate the karst conduits with an A* algorithm ([Hart et al., 1968](#)), which is a shortest-path algorithm similar to the Dijkstra algorithm with an added heuristic. In this pseudo-genetic method, aliasing effects are alleviated by the use of a pipe network conformal to heterogeneities. Starting from the analogy between branchwork karst networks and hydrological networks, they imposed secondary conduits to respect some Horton-Strahler parameters. In the application, the iterative approach was only applied to branchwork networks.

Recently, in the frame of virtual landscape modeling, [Paris et al. \(2021\)](#) proposed a method that reproduces different patterns in 3D, and is flexible enough to permit modifications through an anisotropic cost function. Yet, the reproduction of specific patterns remains difficult and requires tedious calibration steps through manual placement of waypoints. Its application has been limited to small examples without explicit geological constraints or information.

Here, we present a geologically-driven integrative pseudo-genetic approach that reproduces various karst patterns through the use of a cost function. Our main goals include :

- allowing simple control over the morphology of the generated networks with the use of a weighted cost function. In particular, we manage the impact of vadose and phreatic compartmentalization of the system.
- accounting for the uncertainties on the input data—including connectivity information and hierarchy between the inlets—and on the simulator, and propagating them to the final networks in a stochastic workflow.
- Including successive phases of speleogenesis with varying geological and hydrogeological boundary condi-

tions.

To do so, we develop a stochastic method inspired by Paris et al. (2021), enhanced with new functionalities, in order to improve the consistency of the simulated networks. We conduct sensitivity tests on the various input parameters to better understand how they control the karst network patterns. These tests are applied on a simple, but realistic, 3D synthetic model originally proposed by Fandel et al. (2022). Finally, the ability of the method to produce specific karst patterns is assessed by applying it to a real case: the Ribeaucourt karst system, Meuse, France. It corresponds to a complex undercover karst, which has been the subject of studies for dozens of years (e.g., Jaillet, 1995; Devos, 1996). Yet, knowledge on the position and geometry of the karst conduits is still very limited (Jaquet et al., 2004), and a better idea of the characteristics of the karst network would enhance the hydrogeological model and better reproduce hydraulic head and flow rate timeseries.

1 METHODOLOGY

We present the methodology used to simulate discrete karst networks. It is based on the method proposed in Paris et al. (2021), with several modifications accounting for a wider range of data, geological knowledge and hydrogeological contexts. The resulting method is called KarstNSim and is implemented in an open-source C++ code (https://github.com/ring-team/KarstNSim_Public). A general workflow of the KarstNSim methodology is shown in Figure 1.

1.1 Domain sampling

The first step of the algorithm consists in discretizing the domain onto which the following computing steps are carried out. Paris et al. (2021) randomly sample points inside the domain. Connecting those points creates a graph, which can be seen as an unstructured grid. The random sampling uses a uniform Poisson-sphere parameter, r , analogous to a sphere diameter. It corresponds to a minimum distance between each pair of sampled points and can be seen as inversely proportional to the density of the sampled point cloud.

We normalized the parameter r over the dimensions of the domain so that its value range from zero to one. This avoids resolution issues in the z axis due to the usual low depth of the models compared to their x/y extension. With such a discretization type, our method avoids most of the aliasing effects associated with structured grids.

We also propose to use a varying density of sampling points in the domain, according to the geological nature of layers and the presence of heterogeneities. That way, heterogeneities can be better represented. In less relevant areas, such as aquiclude formations, using a lower resolution will speed up computations. To do so, we compute the sampling cloud with a fast point sampling algorithm adapted from the one described in Dwork et al. (2021) for medical imaging, itself adapted from Bridson (2007). The algorithm samples 3D Poisson-spheres with variable radius—and thus, variable density—by scanning information stored in a background grid. This procedure is analogous to a random point sampling with a minimum distance parameter (Figure 2).

To facilitate the exact integration of karst inception features like transmissive faults, inception horizons or water tables, we

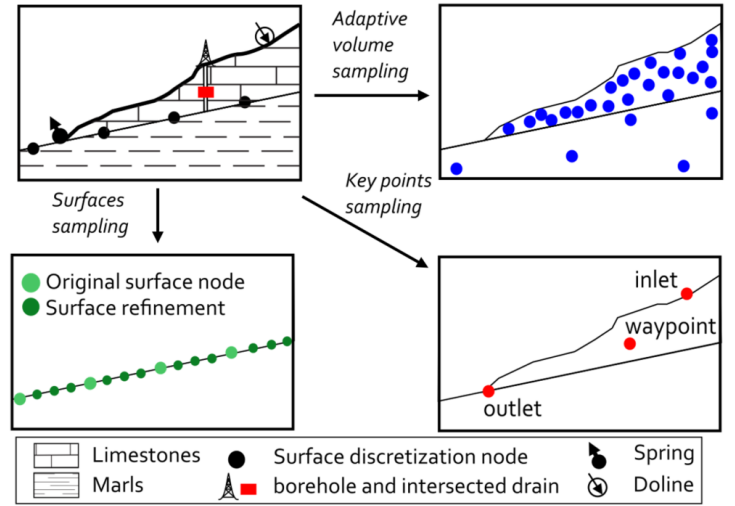


Figure 2 Adaptive sampling of the geological domain. Volumes are sampled with the use of a background grid where local densities are defined. The algorithm used is the one described in Dwork et al. (2021) adapted to 3D. Geological surfaces (faults, horizons, water tables) are sampled by using their internal mesh: vertices of their triangles and mid-points of their edges are added to the point cloud. The red dots correspond to key points, or mandatory pathways for the network.

represent them explicitly with Triangular Irregular Network surfaces (TINs), as usually done in geomodelers (Caumon et al., 2009). The vertices of their triangles are added to the sampling cloud, and the density can be increased by adding mid-points at each triangle edge, creating dyadic triangles. To guarantee that the triangles do not introduce aliasing to the sampling cloud, we resample the mesh with a relaxation process to obtain nearly equilateral triangles.

Key points can also be added to the sampling cloud, corresponding to inlets and outlets of the network, or passages where caves are known or supposed to exist. During the simulation, conduits are constrained to cross these key points.

1.2 Creation of the n-nearest neighbor graph

The second step of the algorithm consists in generating the n-nearest neighbor graph, which edges are used to generate the karst network. The edges are created from each previously generated point towards its N_{nng} nearest neighbors, forming a directed graph. Additionally, an edge is only created if its length is smaller than a value L_{nng} defined by the user. It enables the graph to keep a resolution of at least L_{nng} . Compared to the original version of Paris et al. (2021), we additionally normalized the value of L_{nng} by the largest dimension of the domain, in order to keep an acceptable resolution in every dimension.

1.3 Computation of costs for each edge

A cost is computed on each oriented edge \mathbf{ab} of the graph. This cost can be seen as a local resistance of the medium to karstification: the greater it is and the less likely karstification is to occur in the surroundings of the associated edge. It is designed to encompass the different hydrological and geological factors controlling karstification, but it does not model the processes themselves.

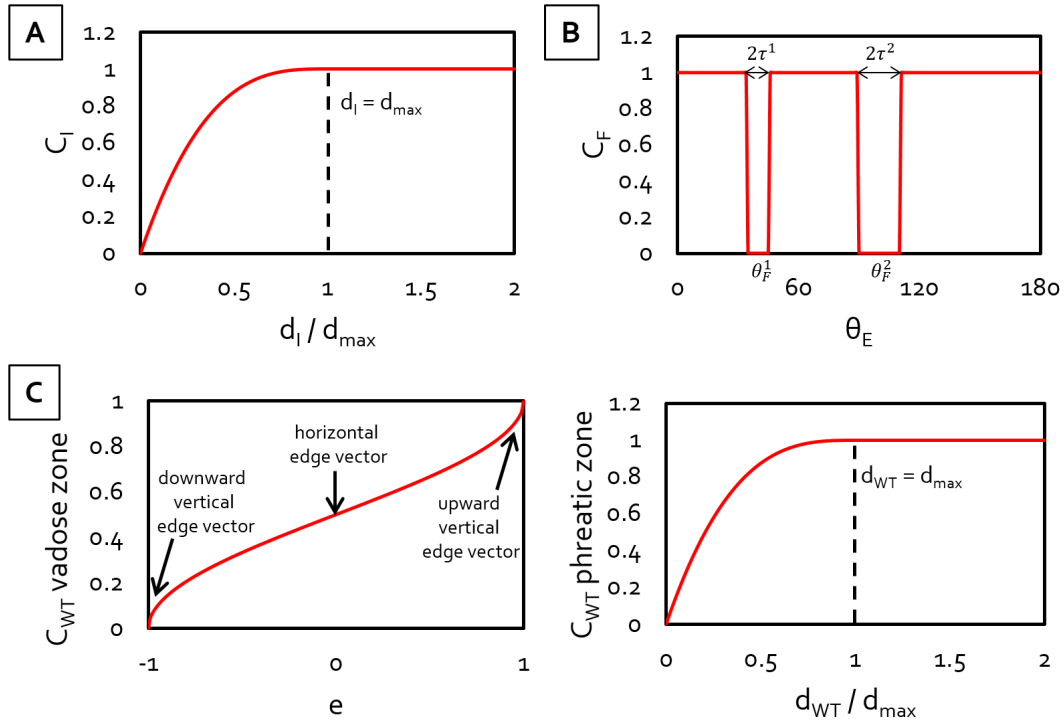


Figure 3 A: Evolution of the inception cost C_I with the normalized distance of the edge to the closest inception surface d_l/d_{\max} . B: Evolution of the fracture cost C_F with the orientation of the edge θ_E , given with the example of two fracture families, each with an average angle θ_F^i and an angular tolerance τ^i . C: Evolution of the water table cost C_{WT} in the vadose and phreatic zones, as function of the scalar product of the upward vertical standard basis vector and the normalized vector of the edge e with the normalized distance of the edge to the water table d_{WT}/d_{\max} respectively.

Compared to previous works, we propose to redefine the cost equation as a weighted combination of normalized terms accounting for the geological and hydrogeological characteristics of the domain, multiplied by the length of each edge. The weighted sum assumes linearity between the cost function (full karstification process) and its subterms (individual physical phenomena). This assumption allows for an easy control of the influence of each subterm on the overall cost. The cost function is given by:

$$C = L(\alpha_I C_I + \alpha_F C_F + \alpha_{WT} C_{WT} + \alpha_P C_P) \quad (1)$$

The parameter L represents the length of the edge, and its multiplication with the other cost components ensures that the cost function remains consistently proportional to the traveled distance. This feature is crucial in regions characterized by graph density variations.

Let $s(x)$ denote a modified smooth step function inspired from Wyvill's falloff function (Wyvill et al., 1999) :

$$s(x) = \begin{cases} 1 - (1 - x/d_{\max})^3 & \text{if } 0 < x < d_{\max}, \\ 1 & \text{if } x \geq d_{\max}. \end{cases} \quad (2)$$

with d_{\max} the maximal distance at which geological or hydrogeological surfaces have an influence on karstification. The cost $C_I = s(d_l)$ represents the influence of the inception surfaces (structural faults and horizons), with d_l the distance to the closest of those surfaces (Figure 3A).

The cost C_F represents the influence of fractures. Fractures are represented implicitly in the method by adding a cost penalty for graph edges, which are not oriented like the main fracture families. By defining a main orientation for each fracture family θ_F^i and an angular tolerance for the family τ^i , and knowing

the orientation of the edge θ_E , the cost is given by the step function:

$$C_F = \begin{cases} 0 & \text{if } |\theta_E - \theta_F^i| < \tau^i \text{ for any fracture family } i, \\ 1 & \text{otherwise.} \end{cases} \quad (3)$$

The cost is therefore minimal and constant while the orientation of the edge is within the tolerance threshold (Figure 3B). We use this step function for the cost instead of a continuous one because fractures are discrete objects.

The variable C_{WT} represents the influence of the water table. With this new parameter, KarstNSim differentiates vadose and phreatic zones. A water table can be specified for each spring, necessitating the definition of a cost for each distinct water table. Its computation depends on the edge \mathbf{ab} (Figure 3C):

- If \mathbf{ab} is located in the vadose zone, the cost is given by $C_{WT} = 1 - \arccos(e)/\pi$, with $e = \mathbf{z} \cdot \mathbf{ab}$ the scalar product of the upward vertical standard basis vector and the normalized vector of the edge. The cost is null if the vector of the edge is pointing vertically downward. It increases by pointing in different directions, with a maximal cost of 1 associated with vertically upward pointing vectors and 0.5 for horizontal vectors. It mimics gravitary control of karstification in the unsaturated zone.
- In the phreatic zone, the cost is given by $C_{WT} = s(d_{WT})$, with d_{WT} the distance between the middle point of the edge and the water table. A minimal cost is hence obtained close to the water table and it gradually increases as the edge is far below the water table.

The cost C_P denotes the Intrinsic karstification potential P of the rock, given by $C_P = 1 - P$. This potential represents

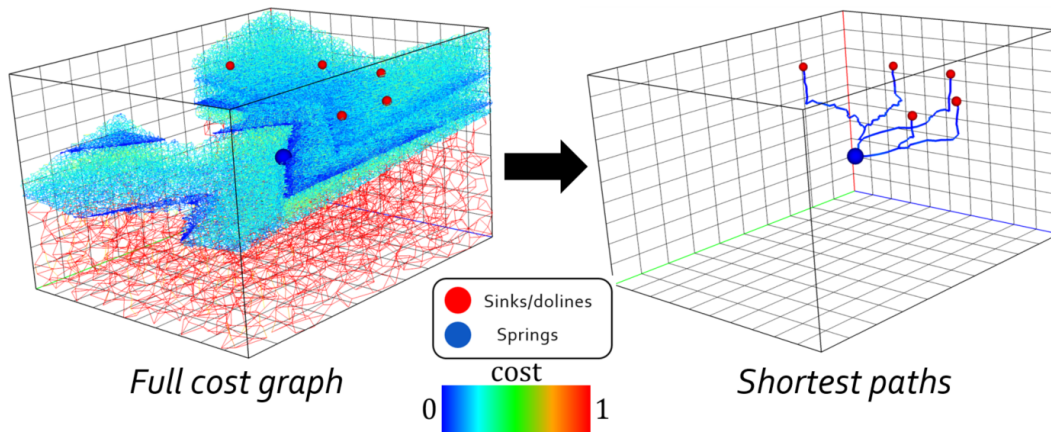


Figure 4 Pruning of the cost graph with Dijkstra's algorithm to create a karst network.

the intrinsic parameters of the geological reservoir which can facilitate water percolation and the dissolution of carbonates (permeability, purity in calcite/dolomite), but does not include the physico-chemical characteristics of water and the nature of recharge. It is expressed between 0 and 1, those two values representing respectively the least and most karstifiable layers potential.

The parameters α_I , α_F , α_{WT} and α_p designate the weights, defined for each cost component.

1.4 The shortest path computation between inlets and outlets

The last part of the algorithm consists in using a shortest path algorithm —here the Dijkstra algorithm— to compute the shortest path in the graph between pairs of inlets and outlets, and with waypoints if those were added. We use a Dijkstra algorithm instead of an A* one because constructing an admissible heuristic —as required in the A* algorithm— in a non-Euclidean space is non-trivial, and does not improve the overall performance. The process generates a karst network between one or more inlets and one or more outlets (Figure 4).

- **Vadose and phreatic zonation**

Vadose and phreatic zones in the aquifer are separated by the water table, linked to one or multiple springs. This surface is explicitly defined in KarstNSim as a Triangular Irregular Network (TIN) surface. It can be assimilated to a horizontal plane at the altitude of its corresponding spring if the hydraulic gradient is negligible, or a plane with a dip equal to the hydraulic gradient if not. It can also take more complex shapes if interpolated from available monitoring well data. This zonation is considered in steady state during baseflow.

The geometries of conduits in the vadose and saturated zones diverge because their main driving mechanism is different : gravity and hydraulic gradient respectively. Usually, vadose conduits display a vertical trend (unless they develop along an inception surface) and phreatic conduits a horizontal orientation (Palmer, 2003). Also, the structure of conduits in the vadose zone is different from the one in the phreatic zone: occasionally, vadose conduits develop independently, while phreatic conduits are generally hierarchically organized to convey water to the outlet, following the hydraulic gradient.

For all those reasons, KarstNSim separates the generation of drains in vadose and phreatic zones, similarly to Borghi et al.

(2012), by running the shortest path algorithm twice with two different targets: first between the inlets and the water table, and second between the vadose ending point and the spring (Figure 5). Two cases can be distinguished in the vadose zone where gravity is the main driving mechanism: either the conduits reach the water table and develop mainly vertically (conduit A in Figure 5), or they first reach the basement of the aquifer where they follow the slope of the substratum to the phreatic zone (conduit B in Figure 5). Since the water table is represented by a surface limited by the extension of the aquifer, the separation of vadose and phreatic conduit generation is enough to respect both those trends. If multiple springs are involved, a different path is generated in both zones, since the water table will potentially be different for each spring.

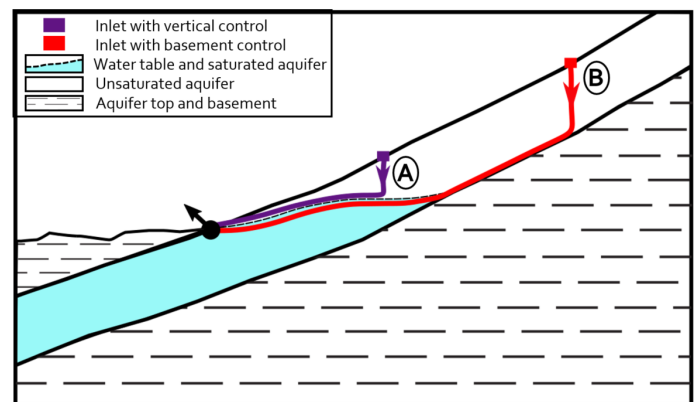


Figure 5 Cross section in a simple karst aquifer, with two karst conduits with different contexts in the vadose zone: one with vertical control (in purple) and one with basement control (in red). The non-planar water table is also represented with a dashed line.

- **Network cohesion control, connectivity matrix and polygenic karsts**

By default, the iteration order for each pair of key points is randomized. Inspired by the iterative approach of Borghi et al. (2012), KarstNSim reduces the costs of edges previously traversed by conduits by multiplying them by a given percentage P_{red} . A P_{red} value of 100% means no reduction in cost, and the degree of cost reduction increases as P_{red} decreases. It implies that paths generated at each iteration influence the next iterations (Figure 6), a feature not present in Paris et al. (2021). With this additional parameter, it is now possible to regulate the network cohesion level through the parameter P_{red} .

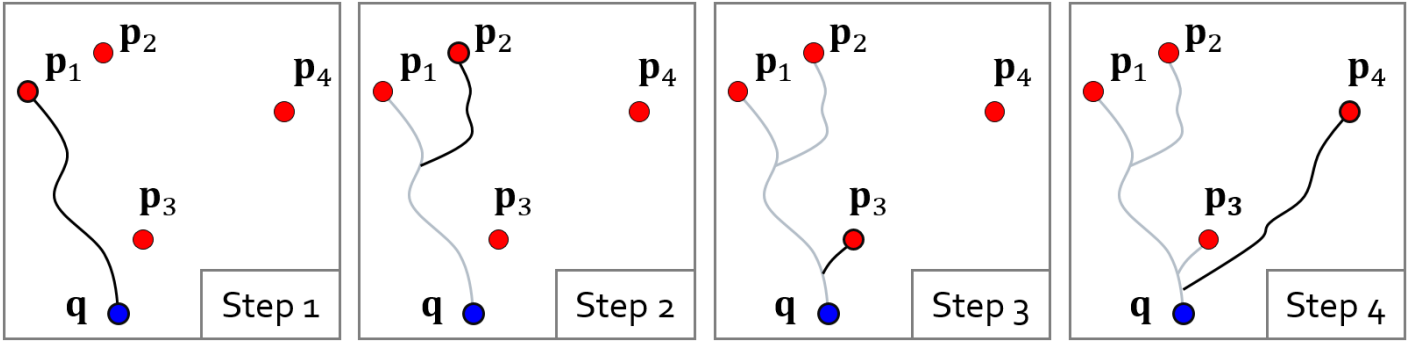


Figure 6 The iteration order for the inlets (p_1 then p_2 , p_3 and p_4) controls the cohesion and the ramifications of the network towards the spring (q). At each iteration, the used paths have their cost reduced by a percentage P_{red} for next iterations, which attracts other branches generated at the next iterations.

Another novelty in KarstNSim is the use of a connectivity matrix denoted as M to define connection between inlets and outlets in cases involving multiple of them. This *a priori* information can, indeed, be deduced from tracer tests or other field investigations. Each column of M represents a spring and each line represents an inlet. Three values can be defined for each couple p and q of inlet and outlet:

- $M_{p,q} = 0$ if the user knows there is no connection between p and q .
- $M_{p,q} = 1$ if the user knows there is a connection between p and q .
- $M_{p,q} = 2$ if the connection remains uncertain. If multiple twos are present in a single line (*ie.* data does not allow determining which outlet is connected to p), a shortest path is computed between the inlet and each possible outlet. Cumulative lengths of each path are compared and only the shortest one is kept.

Finally, we introduce the possibility to mimic polygenic karstification as an iterative process by using output networks of a simulation as inputs to the next one, and by reducing the cost of previously used edges by P_{red} . That way, one can simulate phenomena as different as epiphreatic conduits between low and high water tables, relict karsts, paleodrainage or diffuence between multiple springs.

2 EXAMPLES ON 3D SYNTHETIC CASES

2.1 Model description

We used a synthetic model inspired by the one proposed by Fandel et al. (2022), and modified to be three-dimensional (Figure 7). It is useful for demonstrating the method capabilities without losing too much realism. We added several geological constraints to demonstrate the wide variety of cases that KarstNSim can reproduce. It includes three geological layers : limestones, corresponding to the karstified aquifer, marly limestones corresponding to the substratum of the aquifer and supposed to be poorly karstifiable and a granite intrusion where karstification should be absent. Conduits were allowed but discouraged to develop in the marly limestones. The geological model was discretized into a structured grid of cubic cells of size 10 m, with 100, 80 and 60 cells in the x , y and z directions respectively. Five inlets were set in the upstream part of the model and two springs, labeled S1 and S2, in the downstream

part. While S1 is associated with an unconfined aquifer with the water table in the limestones, S2 is located at the interface between the topography, the limestones and their substratum, featuring a water table of limited extension or none at all. Both water tables are associated with a hydraulic gradient of 4% in direction of the spring, exaggerated on purpose to better see its impact on the simulations. The following tests only take into account the spring S1, except otherwise noted.

Properties for the sampling density and the intrinsic karstification potential were defined on the cells of the background grid (Table 1). A lower value for r was associated with the limestones where most karstification is supposed to occur, and a larger value to the marly limestones. No points were sampled in the granite intrusion: it avoids generating any conduit in that part of the domain, without having to actually constrain the cost function. Similarly, the intrinsic karstification potential (P) was set to be maximal in the limestones and smaller (but not null) in the marly limestones. The resulting point cloud is visible in Figure 7.

The number of neighbors N_{nng} of the graphs generated on this synthetic model was set to 40 (except in Section 2.2), which offers a wide variety of directions available at each node. The maximal length L_{nng} of the graph edges was set to 50 m.

Table 1 Properties of the synthetic model for each geological layer. The Poisson-sphere parameter r corresponds to the average distance between two randomly sampled points, normalized by the domain size, and therefore expressed between zero and one. Intrinsic karstification potential (P) is a simplified representation of the intrinsic parameters of the reservoir rock which can facilitate water percolation and the dissolution of carbonates, and also expressed between zero and one.

	r	P
	[-]	[-]
Limestones	0.007	1
Marly limestones	0.03	0.5
Granite	/	0

2.2 Impacts of the adaptive random sampling on simulations

We aim to explore the theoretical capacity of the random graph to exhibit high multidirectionality in relation to its resolution, allowed by the adaptive random sampling algorithm used in

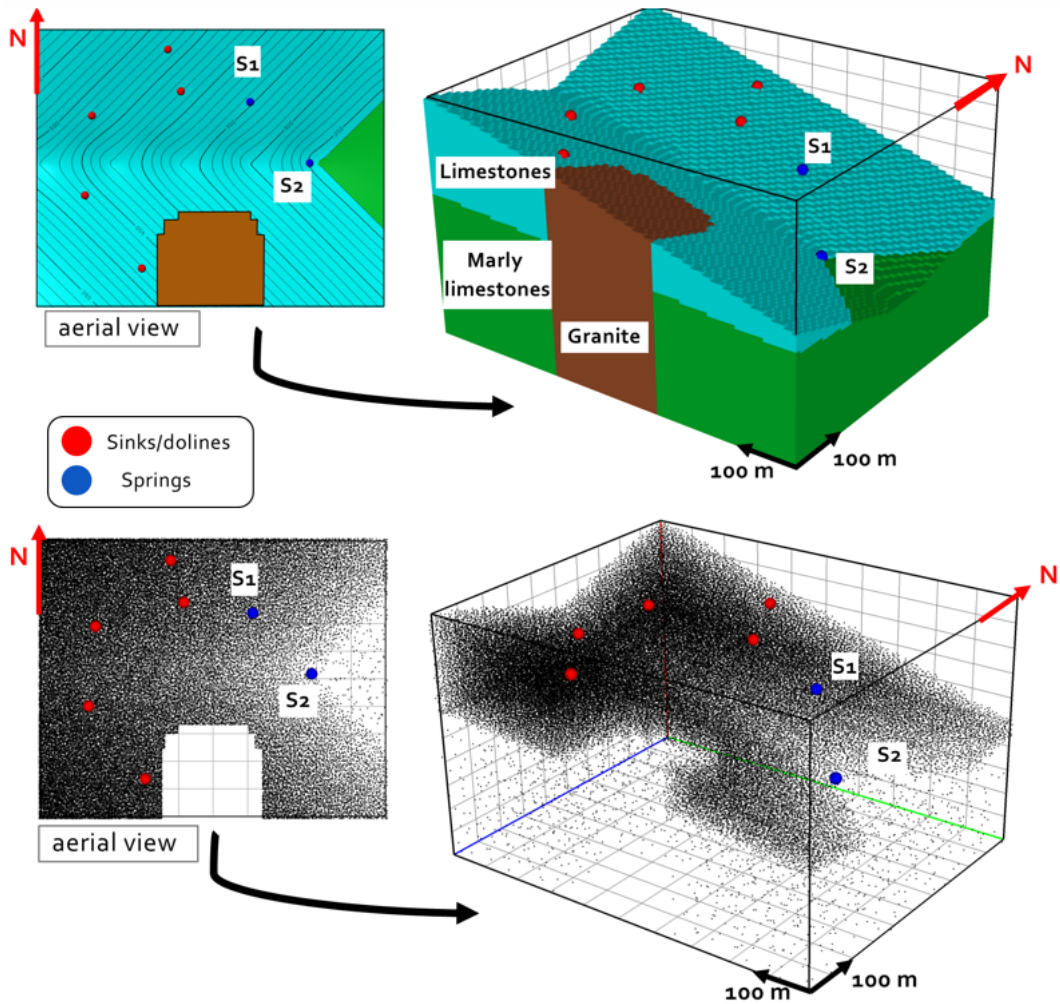


Figure 7 3D geological model inspired by Fandel et al. (2022). Top: the model includes three different layers: limestones in blue, marly limestones in green and granite in brown, as well as five inlets in red and two springs labeled S1 and S2 in blue. Bottom: Corresponding sampling cloud, generated with the adaptive sampling algorithm of Dwork et al. (2021)

KarstNSim (Figure 8).

To do that, we generated networks on two different graphs: a structured, rectilinear grid and an n-nearest neighbor graph based on random point sampling (Figure 9). Both grids have the same number of nodes (85,000), of edges (2,200,000) and of neighbors for each node (26, *i.e.*, one in every direction in 3D). In the case of a homogeneous medium, the network generated in a structured grid is angular and follows the limited directions of the grid, whether it is in the vadose zone—where conduits are vertical—or in the phreatic zone, where an aliasing effect is produced. In the case of the n-nearest neighbor graph, the shapes of the drains are more complex, diverging slightly from a completely vertical trend in the vadose zone thanks to the randomness of the 26 directions available at each node, effectively increasing the resolution of the graph by limiting the aliasing effects.

Aliasing artifacts in structured grids create wrong orientation trends for the conduits, not aligned with fracturation or other heterogeneities. Increasing the resolution of the grid partially alleviates this problem and is computationally intensive. Figure 9 illustrates a case where fracturation has been added in the medium, in two main directions (N30 and N150) that do not correspond to any of the 26 directions available in the structured grid. The results show that vadose conduits in the structured grid are almost the same as those without fractura-

tion, and phreatic conduits still display strong aliasing effects in different directions from fracturation. Also, due to non-presence of sampling points on the water table, phreatic conduits generate just under the real phreatic level and display a small jump close to the spring, which can be seen as a numerical artifact. In contrast, the n-nearest neighbor graph creates less vertical and more angular conduits, just as expected in a fractured medium. Phreatic conduits are also aligned with the fracture main orientations, and develop onto the water table thanks to the introduction of its inner discretization points in the sampling cloud.

Generating multiple networks with the same input parameters produces slightly different results at each iteration due to the randomness of the generated graph, shown in Figure 10, gathering 50 simulations. The deviation between results can be seen as the impact of the discretization on the final conduit position uncertainty. This stochastic workflow therefore enables us to quantify the impact on the uncertainty of the final model set. Such uncertainty propagation can be harder to assess in methods with a deterministic discretization support, where a sensitivity analysis on grid orientation and resolution should be carried out.

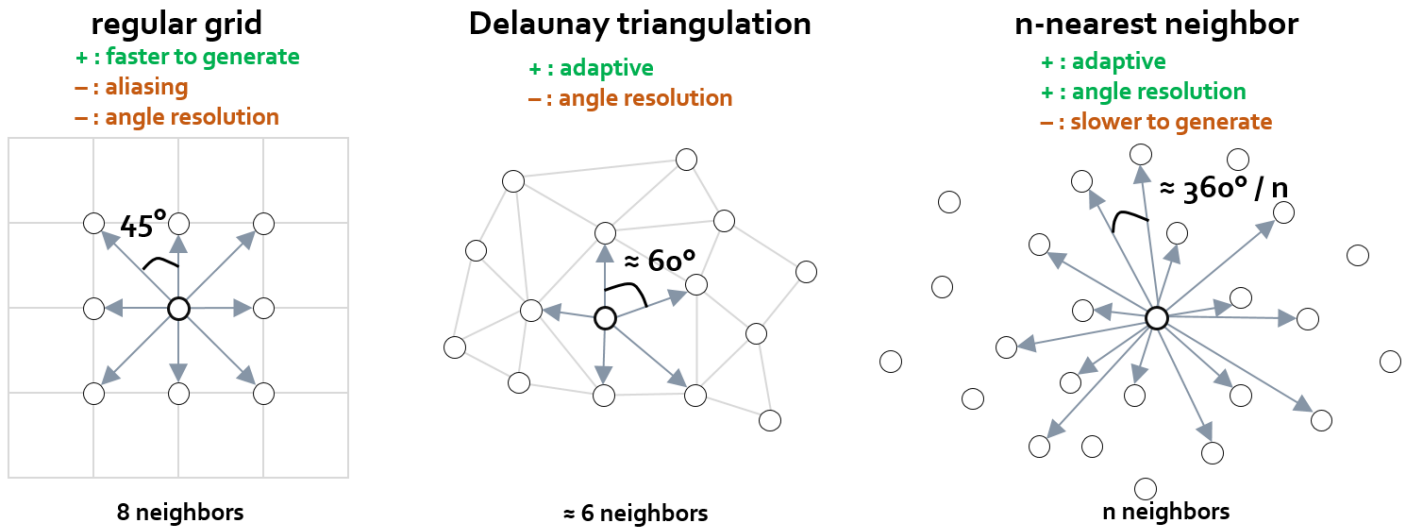


Figure 8 Comparison in graph resolution between regular grid, Delaunay triangulation and n-nearest neighbor graph. Number of neighbors and average angles are given in a 2D case.

2.3 Influence of the number of neighbors

Three main parameters control the characteristics of the graph : the normalized Poisson-sphere property r , the number of nearest neighbors N_{nng} and optionally a maximal edge length L_{nng} . To evaluate the influence of N_{nng} on the level of precision of the discretization, we simulated 10 networks under simplified assumptions: (i) the cost function is influenced by the Euclidean distance only, (ii) a high, constant density is im-

posed with a Poisson-sphere parameter $r = 0.005$ in the whole domain (even in the granite intrusion) and (iii) L_{nng} is supposed to be infinite (*i.e.*, each node in the networks always has N_{nng} neighbors, whatever the distance). This experience was repeated for different values of N_{nng} ranging from 25 to 400. By comparing the cumulative conduit distance of each simulated network with an optimal, straight pathway (vertical in the vadose zone and horizontal in the phreatic zone), we are able to quantify the simulation precision according to N_{nng}

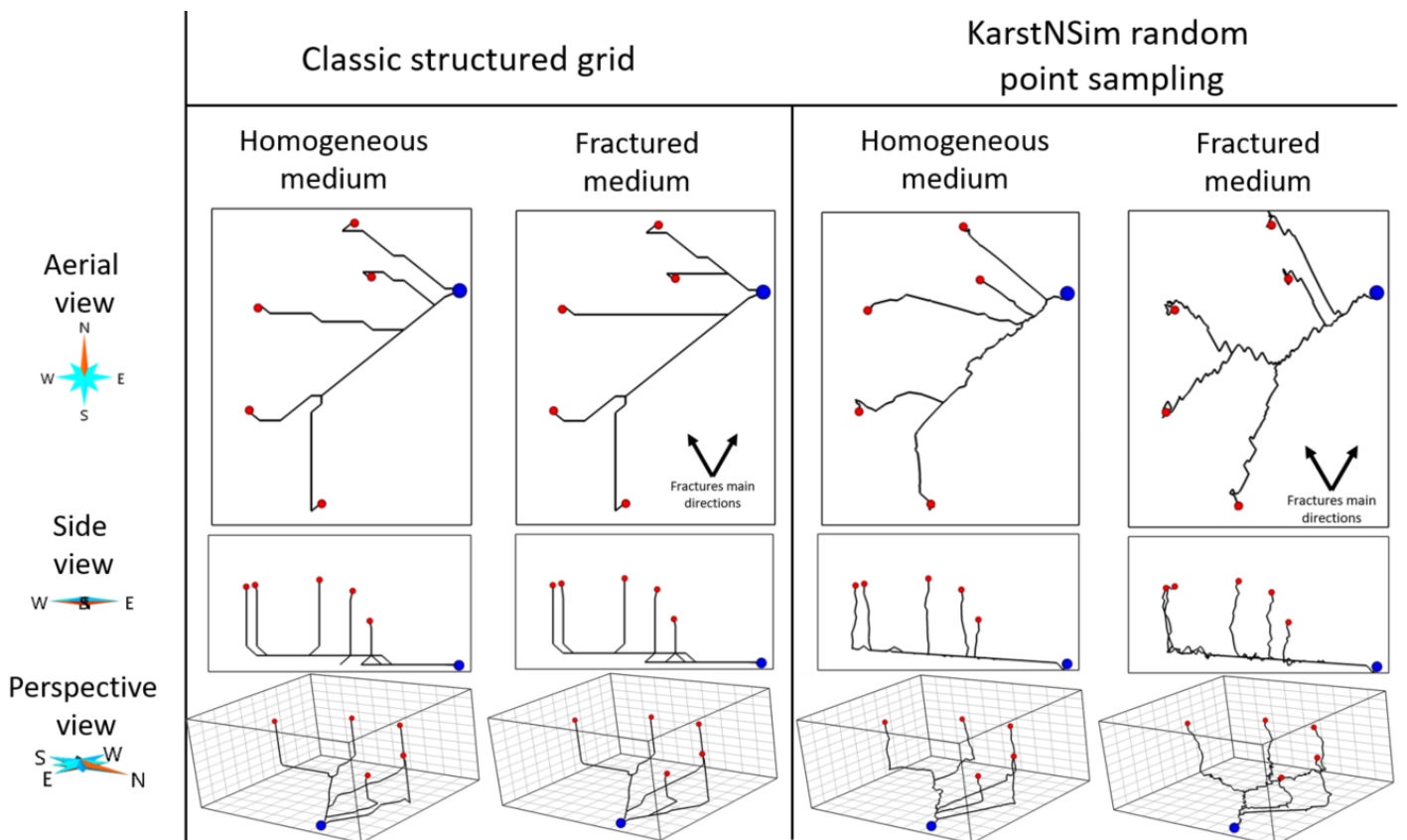


Figure 9 Simulations of the network on a structured grid (left) and on a random point nearest neighbor graph (right) and in two different media: homogeneous or fractured (with two directions : N030 and N150, both with the same weight). Both meshes have the same number of nodes/cells (85,000), of edges (2,200,000) and of neighbors for each node/cell (26). Top: Aerial view. Middle: side view. Bottom : perspective view.

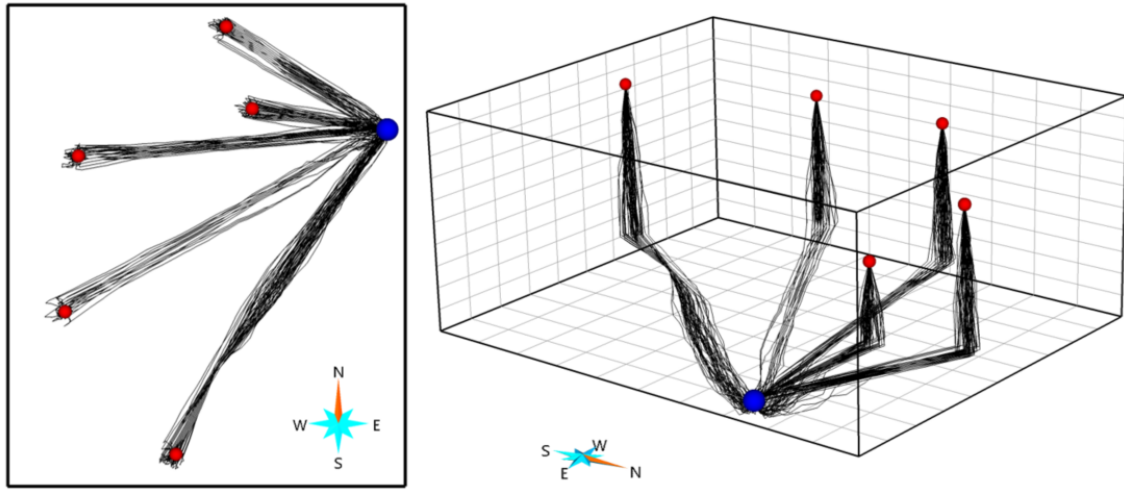


Figure 10 The result of 50 simulations starting from five inlets to the spring S1, only taking into account the water table component of the cost function. The originality of medium discretization proposed in KarstNSim allows to assess part of the uncertainty on conduit location.

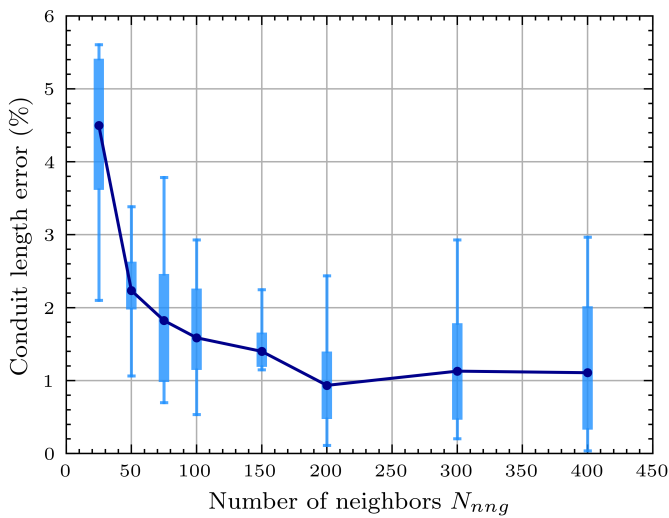


Figure 11 Evolution of the cumulative conduit length error (in %) between simulated conduits and a theoretical trajectory with the number of neighbors N_{nng} for 10 simulations. The blue line represents the average values from the 10 simulations.

(Figure 11).

The cumulative conduit distance tends to decrease with the number of neighbors, until reaching a threshold starting from 100 neighbors with an error of around 1%. The sharpest decrease occurs when increasing N_{nng} from 25 to 75.

2.4 Impact of costs on conduits geometry

2.4.1 Reproducing vadose and phreatic trend separation

A network was generated by using a cost reduction factor P_{red} of 10 % (Figure 12) to illustrate the two-step generation of the network (vadose and phreatic conduits separately). Three different conditions were considered for the limestones and marly limestones: (a) homogeneous media, (b) fractured media with two fracture family directions (N00 and N60) and an angular tolerance of 5° for both families, and (c) a presence of an inception horizon dipping in the direction of the hydraulic gradient, and crossing the domain in the vadose zone. In the case of homogeneous media (Figure 12A), vadose con-

duits develop more or less vertically. Phreatic conduits form more or less horizontally, along the phreatic level. Conduits form in a cohesive way, due to the fixed value $P_{red} = 10\%$. The phreatic level being defined as a triangulated surface, its shape can be interpreted from piezometric data to reproduce more complex hydraulic gradients.

When imposing fractured media in the whole domain (Figure 12B), the vadose cavities are less straight and more angular, even if the gravity effect is still the dominant factor. The same can be said for phreatic conduits, which form stepped, angular patterns aligned with the fractures main orientations. Introducing an inception horizon in the vadose zone, influences drastically the vadose conduit positions (Figure 12C), in the same way as observed on real systems by [Filippini et al. \(2009\)](#). Such inception horizon may represent the top of the substratum that controls underground runoff convergence towards the phreatic zone, or directly to the spring. In that context, vadose conduits reach the water table farther, if not at all. Some conduits even propagate away from the spring, as sometimes observed in the field, because it can be less costly to reach the inception surface as soon as possible.

2.4.2 Influence of implicit fractures

The fracturation density can be controlled by the weight α_F given to the fractures in the cost function (Equation 1). We tested this effect by varying the fracture weight from 0 (no fractures) to 4 (highly fractured medium), and in each case with angular tolerances of 5, 10 and 15° while all other weights were set to 1. Figure 13 shows the results in aerial view to observe the phreatic level conduits. We imposed two fracture orientations, N00 and N60. The more the cost increases, the more angular and rectilinear the network becomes in the phreatic zone, in directions parallel to the fracture orientations. A weight $\alpha_F \geq 2$ tends similarly to produce a very angular network. By increasing the angular tolerance, the networks stay angular but permit more directions to be followed, within the tolerance threshold of the fracture main orientations.

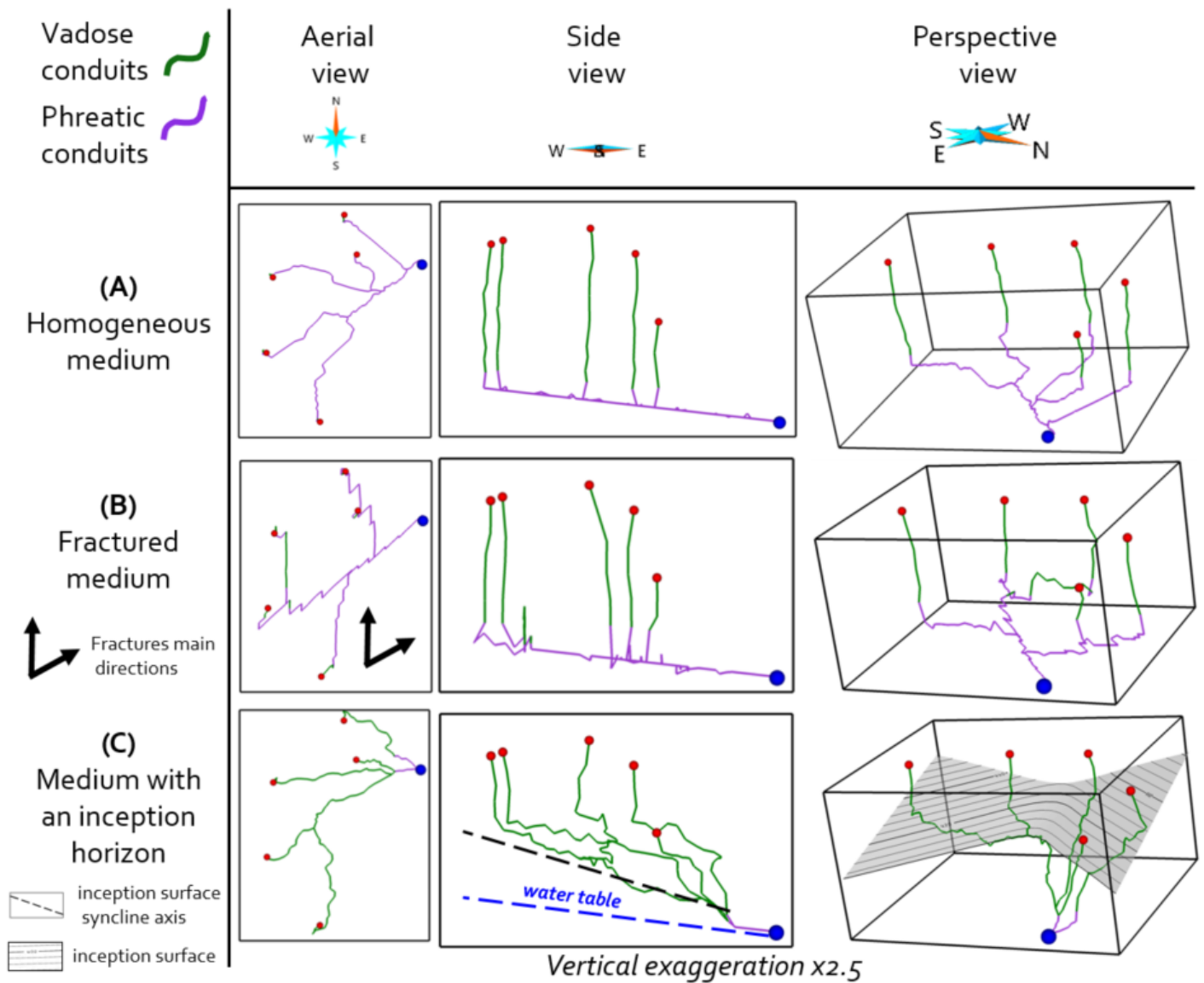


Figure 12 Networks generated in different geological media (in green: vadose conduits, in purple: phreatic conduits). Top : homogeneous medium. Middle: fractured medium with main orientations N00 and N60, both with same weight. Bottom: medium with an inception horizon.

2.5 Control of the cohesion of the network and link with the inlet hierarchy

Conduits being simulated sequentially for each inlet in a fixed or random order, increasing P_{red} makes conduits be more attracted by previously simulated conduits. Therefore P_{red} permits to control the network cohesion level (Figure 14). If the cost of karstified edges is not reduced, the generated network is still slightly ramiform close to the spring, because the cost function naturally attracts conduits and makes them converge towards low-cost valleys.

The cost reduction being applied at each iteration, it is likely that edges crossed by conduits starting from the first inlets (e.g., inlet 1 in Figure 14) at the first iterations have their cost reduced several times, making the corresponding inlets more impactful in the overall network.

2.6 Polygenic processes

If a karst system results from multiple phases of karstification, modeling the corresponding karst network in one step could be difficult. Instead, we propose to simulate parts of the net-

work for each phase of karstification sequentially, with each step influencing the following ones, as originally proposed by Borghi et al. (2012). We simulated both configurations, a base level drop and a base level rise, between the springs S1 and S2 of our synthetic model (Figures 15 and 16 respectively). Each test uses a cost function with all weights set at 1 and with $P_{red} = 0.9$. The cost of edges karstified during step 1 is reduced for step 2 by 50%.

In the case of a base level drop (Figure 15), conduits are first generated from the inlets to the spring S1. Then the simulated karst network is used as initial condition for the second simulation towards the spring S2 at a lower altitude. The second simulation reuses the vadose conduits and part of the phreatic conduits of step 1, but then new conduits drop again until reaching the substratum of the limestones, along which they propagate downward before finally reaching the spring S2. The resulting final network is similar to real networks with abandoned paleodrain linked to a base level drop (e.g., Audra and Palmer, 2013).

To mimic a base level rise (Figure 16), we added S1 as an inlet during step 1, and as an outlet during step 2. The spring S2, added as an outlet during step 1, is plugged during step

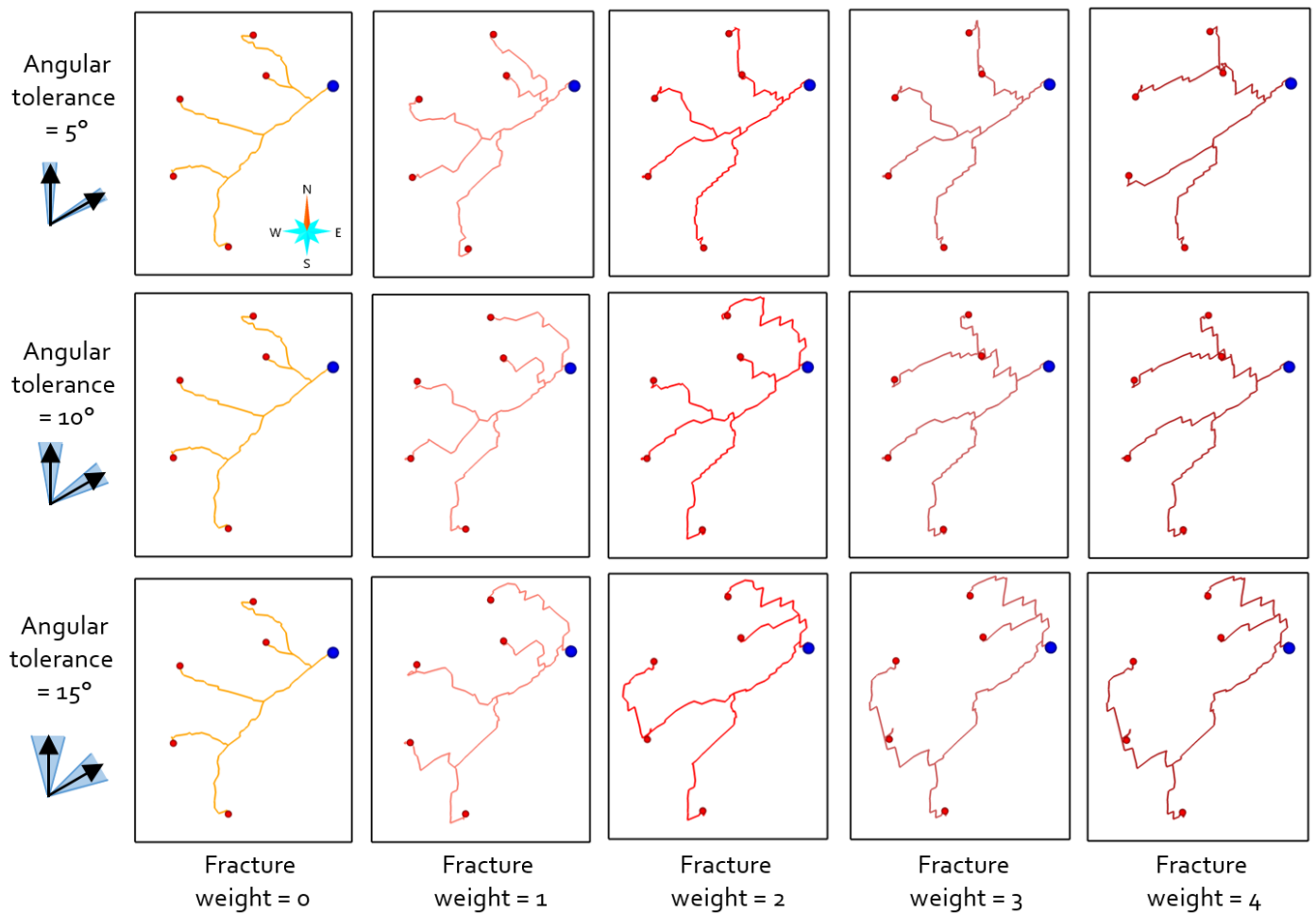


Figure 13 Influence of the weight of implicit fractures in the cost function α_F on the geometry of networks in the phreatic zone.

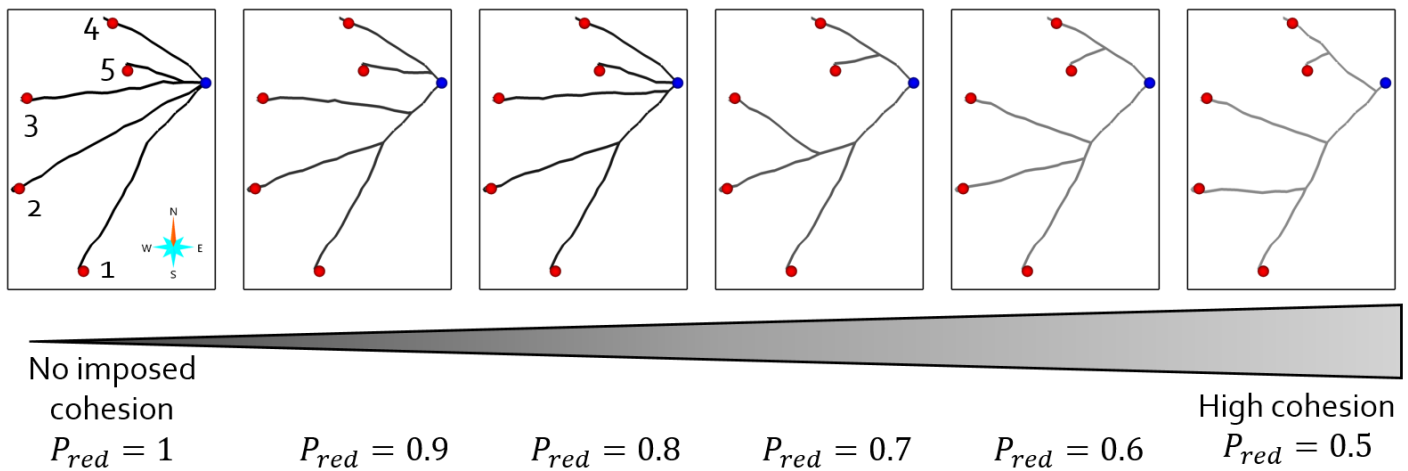


Figure 14 Influence of the parameter P_{red} on the level of cohesion of the network in the phreatic zone. On the left: higher values of P_{red} . On the right: smaller values of P_{red} . The iteration order of inlets is also indicated on the left-most figure.

2 due to clay infilling in the valley. Drains are first generated sub-vertically until reaching the substratum, and then follow it before arriving at S2. In step 2, conduits starting from inlets far from S1 tend to reuse paleodrain at the substratum level, until arriving close to S1, where the sub-vertical drain created from S1 during step 1 is reused, connecting the paleo-conduits to S1, and mimicking the structure of a vauclusian karst system. The drain starting from the inlet close to S1 directly reaches S1 by following the water table, without dropping to the substratum, and hence mimicking a jurassian karst system. Relict

paleodrain remain in the final network, still connected to a now-inactive spring S2. The methodology hence enables modeling realistic polygenic karst networks, composed of inactive drains.

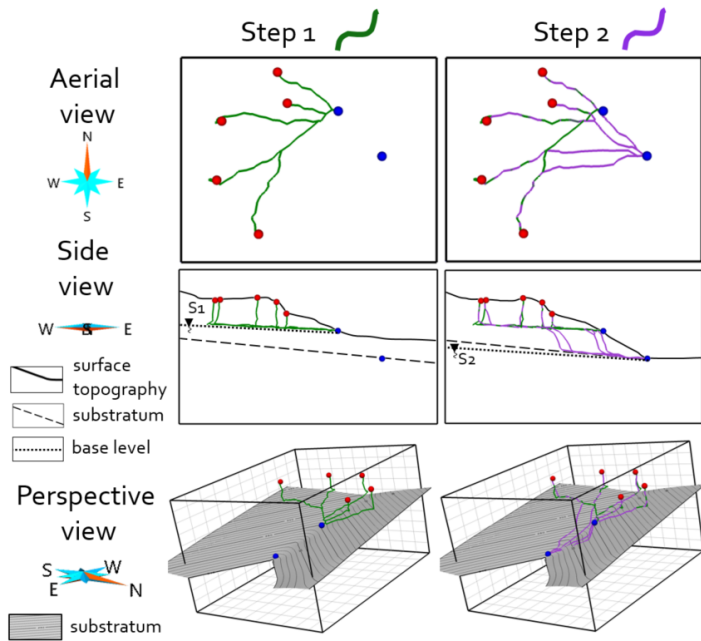


Figure 15 Two-step simulation mimicking a base level drop, seen in aerial, side and perspective view.

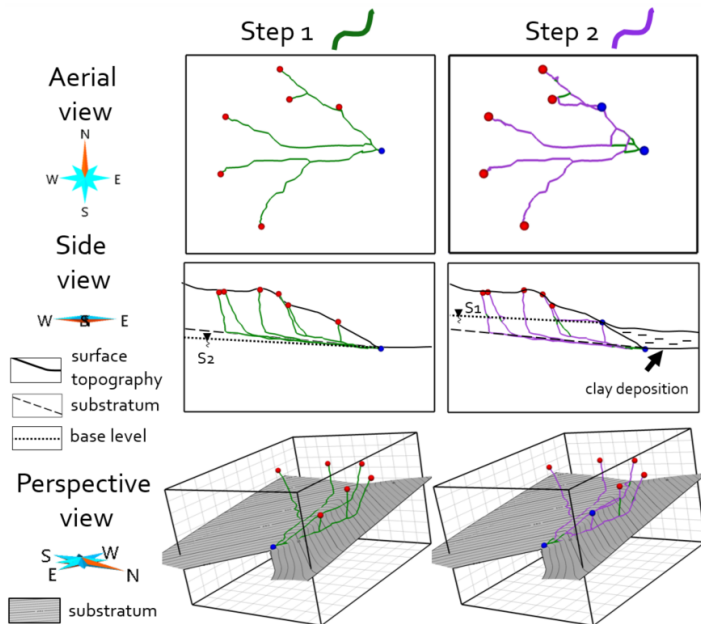


Figure 16 Two-step simulation mimicking a base level rise, seen in aerial, side and perspective view.

3 CASE STUDY : EXAMPLE OF THE BARROIS KARST NETWORK

3.1 Site description

We applied our methodology to the Barrois limestones, which is an extensively studied karstified aquifer formation located in the east of the Paris Basin in Meuse, France (Figure 17A) to show its ability to generate realistic karst networks. Despite the existing studies in the area, the position and geometry of the karst conduits are mostly unknown. The area of the formation being large (more than 1,100 km²) and the geological and hydrogeological contexts being diverse in the region, we choose to focus on a part of the Barrois corresponding to the probable recharge area of the Ribeaucourt spring, which is

the major spring of the southern Barrois plateau. Aerial photographs and Lidar imaging have detected 165 inlets, mostly located on the ridges of the valleys, and clustered in three areas located south, west and east from the spring (Figure 17B). In that part of the domain, four main formations are outcropping (Figure 17C): the Dommartin limestones and Sublithographic limestones (aquifer layers), the *Pierre Châline* marls and Kimmeridgian marls (aquiclude layers). All those formations dip towards the north/northwest (N335).

We have multiple pieces of evidence indicating the presence of a karst system within the limestone levels, and some interbeds of the Kimmeridgian marls, represented in Figure 17C by bold black lines. This evidence encompasses tracer tests, drains intersected by boreholes, presence of many sinking rivers, dolines and springs and high fluctuation of the flow rate in rivers. The spring is outcropping in the *Pierre Châline* formation, close to the interface with the Dommartin limestones through which water is probably drained to the spring. Tracer tests have also shown that sinkholes outcropping in the Sublithographic limestones are connected to the spring, which means that the low-permeability *Pierre Châline* marls are traversed by water, probably through enlarged fracture drains (Andra, 2021). We therefore make the assumption that the Ribeaucourt spring is supplied by two reservoirs: the Dommartin limestones by direct drainage during high water periods and the Sublithographic limestones by indirect drainage upward through the *Pierre Châline* marls during low water periods. The conditions at which both reservoirs, the Dommartin and Sublithographic limestones, are drained, and their relative prevalence in the Ribeaucourt karst system is also an issue that we want to investigate.

We have little information on the real morphology of the network, and thus cannot determine an optimal set of weights for the terms of the cost function because we do not know the driving parameters of karstification for that karst network. We therefore want to generate possible morphologies for the karst network connected to the Ribeaucourt spring through different sets of main driving parameters of karstification. For these different parameter sets, the goal is to evaluate whether the different networks can mimic the structuration of conduits in both the Dommartin and Sublithographic reservoirs, as well as their relative prevalence.

3.2 Simulation settings

The aim is to control the simulation by adjusting the weights of the cost function and the parameter P_{red} . To this end, we propose four distinct scenarios for the morphology of the karst network (Table 2). Each scenario represents a different hypothesis regarding the network morphology. The first test corresponds to a default scenario, where no driving parameter is predominant over the others. We implemented this by setting all cost weights to 1, no cohesion ($P_{red} = 1$) in the vadose zone, and a light cohesion ($P_{red} = 0.8$) in the phreatic zone. The second test represents the hypothesis of a highly fractured medium, where the weight of fractures is tripled compared to the default scenario. The third test illustrates a case where the lithology of the different layers is the principal controlling parameter of karstification. It includes the effect of inception horizons in the form of bedding planes, and the various intrinsic karstification potentials of each layer. The fourth and last test corresponds to a case with an imposed stronger cohesion in the network through the use of a lower value for P_{red} , for both phreatic and

vadose conduits.

A 3D geological model was previously built by ANDRA (French national agency for the management of radioactive waste) with the geomodelling software SKUA-Gocad (Andra, 2021, Figure 18). We reused it here, and added a horizontal surface corresponding to the theoretical water table drained by the Ribeaucourt spring in the Dommartin and Sublithographic limestones, assuming a negligible hydraulic gradient. We defined Poisson-sphere parameters r of 0.01 in the aquifer layers and of 0.1 in the aquiclude layers. We also set an intrinsic karstification potential of 1 in the aquifer layers and 0 in the aquiclude layers. From the initial 165 inlets detected in the surroundings of the spring, we selected randomly and kept 30 of them for the simulation, to reduce the network density and hence simplify the interpretation of the results. Fractures were also added through two main directions, N060 and N153, extracted from rose diagrams of traces of fractures spotted on aerial photographs of the Barrois plateau (Andra, 2021). An angular tolerance of 9° was associated with each fracture family orientation, and the same weight was attributed to each family. For all four tests, the graph parameters N_{nng} and L_{nng} were set to 50 and 500 m respectively, and the maximal distance of influence of surfaces d_{\max} was set to 50 m.

3.3 Results

Our algorithm was implemented in C++ and the models were simulated on a laptop equipped with Intel® Xeon® W-11855M clocked at 3.20 GHz with 32 GB of RAM. The computational time of one simulation of the Barrois karst was 112.5 s on average, including 22 s to generate the sampling cloud, 90 s to create the cost graph, and 0.5 s to compute the network. A large part of the computational time (65 s out of the 90 s needed to compute the cost graph) is due to the operations on the triangulated surfaces (water tables, topography, inception horizons, faults), mainly computing the position of each node relative to those surfaces which could be largely optimized in the future.

The different results are shown using aerial, perspective and side views (Figure 19), the latter being relatively similar to the one used in Figure 17C. In perspective and side view, some inception horizons of Figure 17C are depicted with 3D surfaces. In the default test, the network forms cohesive branches which slightly follow the fracture orientations in aerial view. As seen in perspective view, conduits from the south mostly follow the inception horizons in the vadose zone and plunge before reaching the water table in the Sublithographic limestones, where they form a dense network of sub-horizontal conduits, that can be seen as the drain collectors of the network. Those phreatic conduits are better seen in side view. The collectors traverse the *Pierre Châline* aquiclude until reaching the spring. Perspective view shows that drains from the west and east of the spring develop sub-vertically and with a low cohesion in the Dommartin limestones. On the side view, we can see that after reaching the top of the *Pierre Châline* marls, they cross the marls vertically to connect to the drain collectors in the Sublithographic limestones. Overall, conduits are predominantly found in the more permeable layers, or along the inception horizon surfaces. It demonstrates the influence of the intrinsic karstification potential and the distance to inception surfaces in the cost function. The generated network is consistent with the initial assumption of a spring directly supplied by the Dommartin reservoir, and indirectly (through

the *Pierre Châline* marls) by the Sublithographic limestones reservoir.

In the second test, corresponding to a highly fractured medium, drains form more angular shapes, similarly to the planar fracture orientations. In perspective and side view, the morphology of the network closely resembles the one obtained in the default setting. The primary distinction lies in the increased angularity and tortuosity due to the fractures introducing noise in the cost function. Nevertheless, conduits are still mostly present in the Dommartin and Sublithographic reservoirs which is consistent with the initial assumption of dual reservoir drainage.

In the third test, the network morphology looks quite similar in aerial view to the one obtained in the default setting, but the side view shows that conduits remain preferentially in the more permeable layers, mostly the Sublithographic limestones, and barely propagate in the Kimmeridgian marls. The reason stands in the increased cost associated to lower intrinsic karstification potentials.

Finally, the fourth test displays a network with a higher cohesion, visible both in aerial, perspective and side view, in phreatic and vadose zones. The final network is thus more ramiform than those of the other tests.

With these four tests, we demonstrate the capacity of the method to generate different realistic morphologies of the network with the same input hydrogeological model, by only changing the weights of the sub-terms of the cost function.

3.4 Simulated networks analysis with graph metrics

Assessing the quality of a stochastic simulation result is not an easy task as it requires accurate objective criteria, independent from the input parameters. Collon et al. (2017, 2021) proposed and evaluated 21 geometrical and topological metrics to describe karst networks. They selected 8 of them, more relevant to analyze networks. They also provided the values computed on a database of 31 real systems. We computed the same metrics on 10 simulated Ribeaucourt karst networks for each of the four different test hypotheses, thanks to the Karstnet code (<https://github.com/karstnet/karstnet>). We then compared them to the ones of the database (Table 3).

The topological metrics of the simulated networks globally fall into the range of the ones computed for the real karsts, while it is less the case for the geometrical metrics. Indeed, as already underlined by Collon et al. (2017), geometrical metrics are affected by the measurement methodologies and the scale of the karstic system, which could justify the choice of other authors not to consider these metrics for simulation result assessment (Kanfar and Mukerji, 2023). However, we found interesting to present them and try to link their values to simulation parametrization. The orientation entropy H_o expresses to what extent preferential directions exist for karst conduits. Consistently with what is observed for real systems, H_o decreases in our simulations as the fracturation and inception surface influence increases. However, it is globally lower for the simulated networks (but remains within the observed ranges for three scenarios) than for the real karsts. It could be partly explained by the influence, in our algorithm, of the Euclidean distance between inlet and outlet, which tends to favour segments oriented along this direction, trend also favoured by the high sampling of the domain. The amplification step proposed by Paris et al. (2021), which incorporates a segment refinement with a small displacement of added points to increase

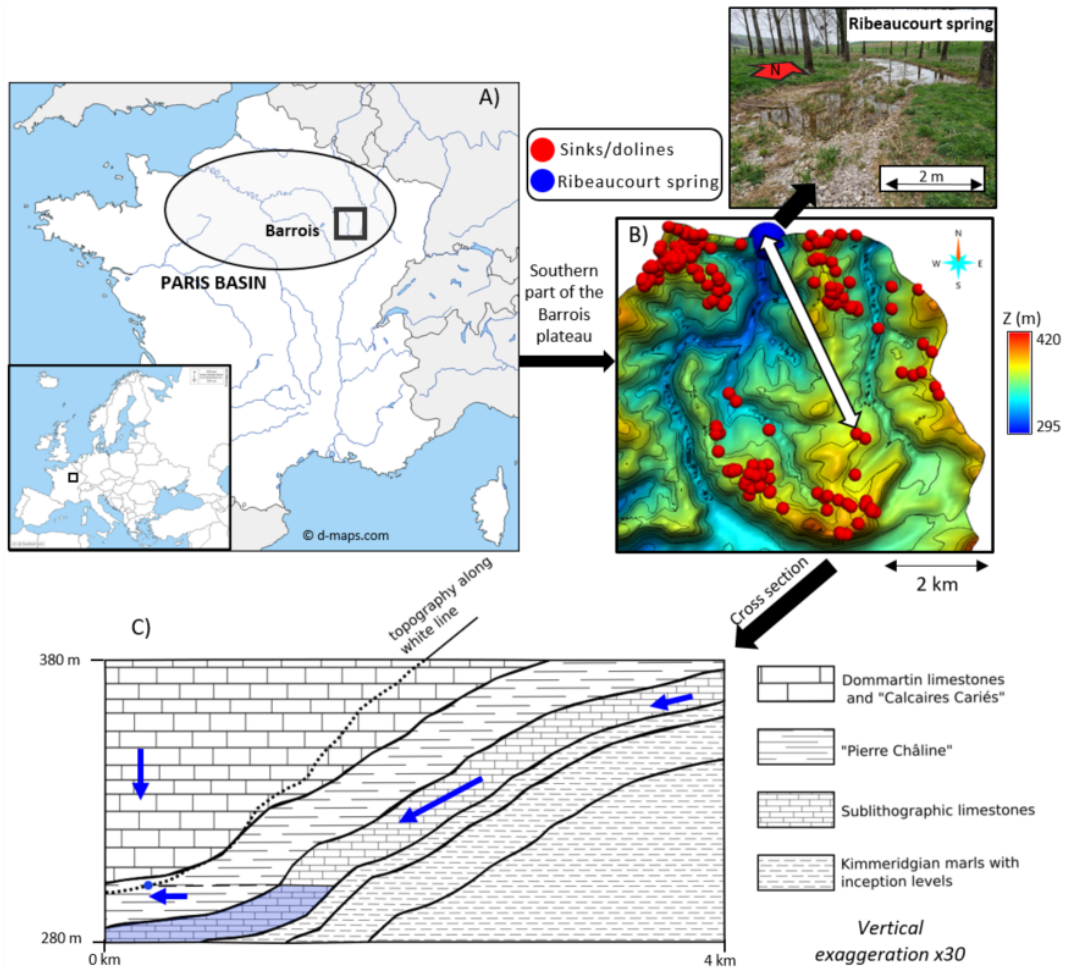


Figure 17 Geographical, geological and hydrogeological context of the Barrois limestones. A) Location of the Barrois plateau in Europe and France. B) Zoom on the southern part of the Barrois plateau. In red: inlets spotted on aerial photographs and Lidar imaging. In blue: Ribeaucourt spring. C) Full cross section of the geological series in the southern region of the Barrois limestones, with visible topography along the cross section represented by a white line in B). The geology above topography was kept to give information on the context in the valley ridges, where some inlets are present. Inception horizons are depicted by bold black lines. The blue area corresponds to the aquifer linked to the spring, represented by a blue point. The blue arrows symbolize probable water flow direction in the two reservoirs.

Table 2 Description of the parameters for each proposed simulation test of the Ribeaucourt karst network.

Tests	water	inception	fracture	intrinsic	P_{red}	P_{red}
	table cost	surfaces	cost	karstification		
	weight α_{WT}	cost weight α_I	weight α_F	potential weight α_P	vadose	phreatic
test 1: default settings	1	1	1	1	1	0.8
test 2: fracture control	1	1	3	1	1	0.8
test 3: lithological control	1	3	1	3	1	0.8
test 4: highly cohesive network	1	1	1	1	0.7	0.7

tortuosity, could be a way to minimize this effect. Concerning the length of network branches \overline{len} , *i.e.* parts between two junction nodes, increasing the cohesion of the network consistently decreases \overline{len} . But the simulated networks display higher values than what is usually observed. Indeed, the Ribeaucourt domain is quite large and we decided to simulate only the main conduits by selecting a subset of the potential inlets (30 over 165). This choice mechanically tends to increase the average branch length as only reaching a junction node, *i.e.* a confluence or a bifurcation between conduits, ends a network branch. Thus, for a given domain size, the less dense a network is, the longer its branches would be. We ran 10 additional simulations within the highly cohesive scenario but with the 165 original

inlets, and obtained on average a mean branch length \overline{len} of 342.29 m, a coefficient of variation of lengths CV_{len} of 1.33 and a length entropy H_{len} of 0.34, all three values being much closer to those found in the database. Again, using the amplification step, could also tend to reduce these discrepancies. Concerning topological parameters, their distribution for real systems and for simulations exhibit large similarities. The average vertex degree \overline{k} is typically below 2 in simulations, whereas it generally hovers around 2.14 in reality. A degree exceeding 2 is indicative of the presence of cycles in the network (Collon et al., 2017). In KarstNSim, the generated networks are predominantly branching, and cycles rarely occur, explaining the low values of the average vertex degree \overline{k} and of the correla-

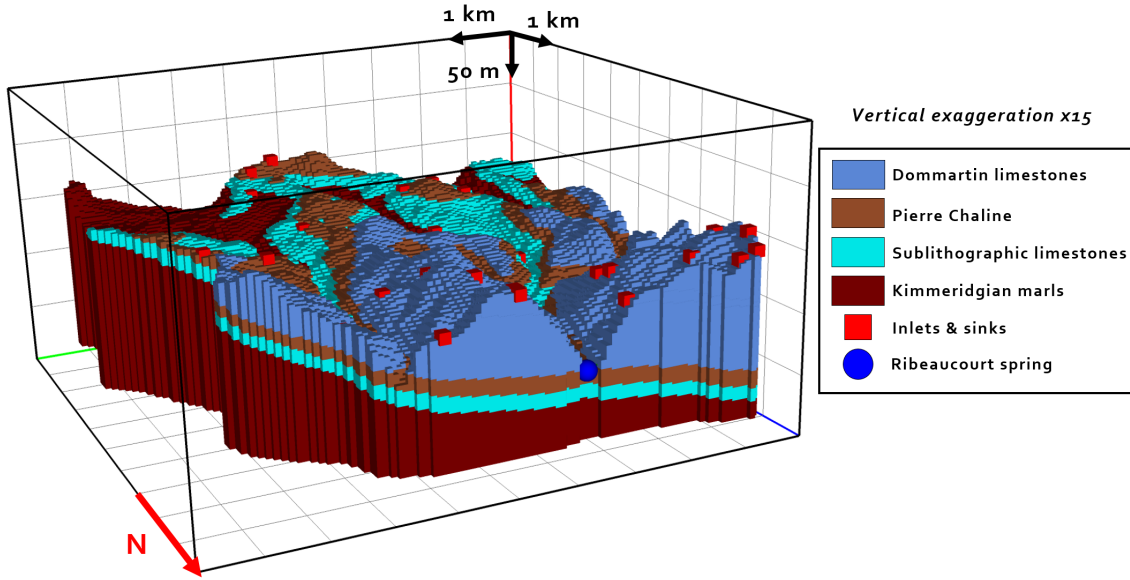


Figure 18 3D hydrogeological model developed on SKUA-Gocad used in this study (Andra, 2021).

Table 3 Minimum and maximum values of geometrical and topological metrics from surveyed cave networks, and minimum, average and maximum values from 10 simulated Ribeaucourt system karst networks for each of the four different test hypotheses: default settings, fracture control, lithological control, and highly cohesive network. Values in bold are within the literature range.

Geometrical metrics				
	Orientation entropy H_o	Average length of branches \overline{len} [m]	Coefficient of variation of lengths $CV_{len}[\%]$	Length entropy of conduits H_{len}
Literature (Collon et al., 2017, 2021)	0.75 / 1.00	8.46 / 331.74	0.90 / 2.56	0.07 / 0.67
1. Default settings	0.78 / 0.80 / 0.83	1023.70 / 1140.00 / 1228.62	1.00 / 1.07 / 1.18	0.63 / 0.72 / 0.77
2. Fracture control	0.62 / 0.63 / 0.65	1066.28 / 1230.36 / 1339.82	0.86 / 1.00 / 1.12	0.67 / 0.73 / 0.81
3. Lithological control	0.86 / 0.89 / 0.91	879.86 / 1023.12 / 1110.73	0.84 / 0.98 / 1.08	0.71 / 0.75 / 0.85
4. Highly cohesive network	0.83 / 0.86 / 0.88	802.72 / 890.38 / 956.17	0.86 / 0.91 / 0.99	0.60 / 0.70 / 0.82

Topological metrics				
	Average vertex degree \bar{k}	Correlation of vertex degrees r_k	Average shortest path length \overline{SPL}	Central point dominance CPD
Literature (Collon et al., 2017, 2021)	1.75 / 2.57	-0.56 / -0.15	2.32 / 55.72	0.02 / 0.57
1. Default settings	1.96 / 2.01 / 2.07	-0.51 / -0.38 / -0.25	6.07 / 6.79 / 7.46	0.37 / 0.48 / 0.62
2. Fracture control	1.96 / 2.00 / 2.07	-0.47 / -0.35 / -0.15	6.16 / 6.77 / 7.40	0.39 / 0.52 / 0.58
3. Lithological control	1.96 / 2.00 / 2.13	-0.53 / -0.40 / -0.29	5.50 / 6.58 / 7.26	0.39 / 0.50 / 0.64
4. Highly cohesive network	1.96 / 1.97 / 2.00	-0.57 / -0.42 / -0.31	6.34 / 6.86 / 7.43	0.43 / 0.53 / 0.57

tion of vertex degrees r_k : high-degree nodes are less common in simulations and tend to be more frequently connected to low-degree nodes.

Overall, the computation of simulated networks morphological metrics shows that simulations are realistic, with most discrepancies explained by the absence of amplification in the network and the choice of reducing the number of inlets for figure clarity.

4 DISCUSSION

Our method is able to generate realistic 3D karst networks in an epigenic context, and allows for easy geology-driven control of the morphology through the different weights of the cost function. It allows for control over the cohesion level, the sharpness of network angles and orientation trends in both the vertical and horizontal directions within the vadose and

phreatic zones respectively.

KarstNSim enables the geologists to tweak the cost function parameters for testing hypotheses and generating a wide variety of networks. The validation or rejection of these networks hinges on surveyed cave maps, or their morphometric properties, as outlined by Collon et al. (2017, 2021). This approach proves particularly efficient in cases where limited data are accessible regarding network morphology. Furthermore, it extends the propagation of uncertainty regarding conduit positions to the final outputs by generating diverse morphologies. Compared to the original algorithm described in Paris et al. (2021), we add the ability to adaptively sample the 3D space of the domain according to geological knowledge. We also propose a complete rewriting of the cost function with new terms (vadose and phreatic trends, intrinsic karstification potential), and normalize and weight all the different costs to simplify their use. The new algorithm also incorporates functional-

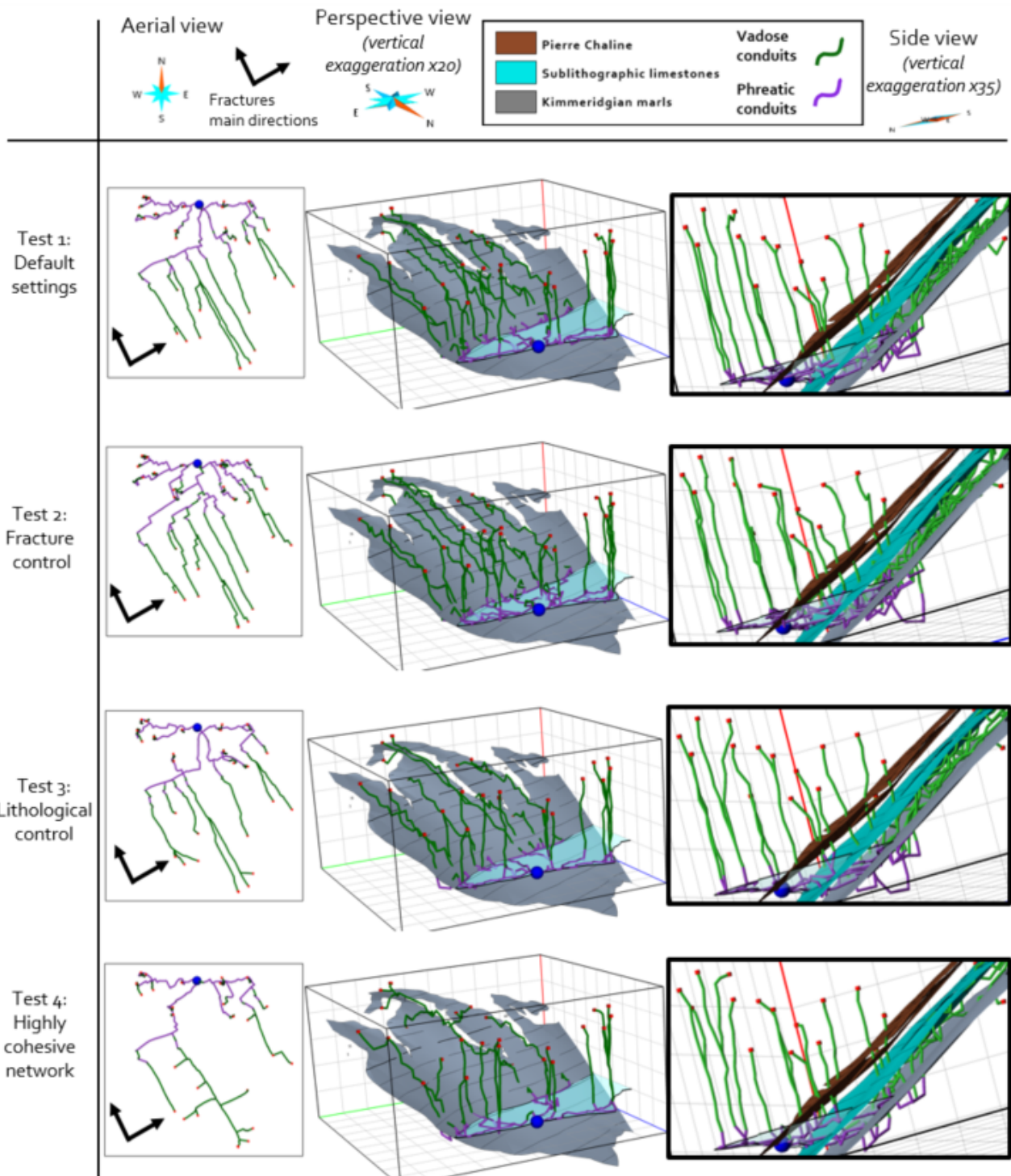


Figure 19 Simulated karst network of the Ribeaucourt spring, under various hypotheses. Networks are shown in aerial view, perspective view and side view. The top surfaces of the Pierre Châline, Sublithographic limestones and Kimmeridgian marls are represented in perspective and side view. Inlets are represented in red and the Ribeaucourt spring with a blue sphere.

ties originally proposed in other methods, such as the sequential simulations for polygenic karsts, and the cohesion factor (Borghì et al., 2012). The following paragraphs aim at discussing the characteristics, assets and limitations of KarstNSim.

4.1 Adaptive random sampling

KarstNSim uses an adaptive sampling of the domain, with a Poisson sphere process presented in Section 1.1. This differs from most methods in the literature, as they typically employ a structured grid as the discretized domain of the simulation. In Borghì et al. (2012), all properties and heterogeneities are

stored in the structured grid, which may involve aliasing effects, especially in the case of a fractured medium. The grid is also not conformal to structural heterogeneities. Collon-Drouaillet et al. (2012) use a pipe network conformal to heterogeneities to alleviate this problem. The grid-less approach proposed in Paris et al. (2021) and reused in our paper also limits the aliasing artifacts of grid-based approaches, as shown in Figure 9, and is fully conformal to heterogeneities. While the conduits generated in Figure 9 exhibit notable differences in appearance for both cases (structured grid and random point sampling graph), it should be noted that, in the context of flow simulation, these variations in network shapes are not

likely to significantly modify the computed spring discharge. Hence, the choice of the domain discretization is mostly dependent on the needs of the end-user, with structured grids being acceptable choices for low resolution, large-scale models and unstructured grids being more appropriate in high resolution and/or small-scale models.

However, varying the density of point sampling in the domain according to the karstification potential of the layers gives a computational advantage compared to a constant density sampling in the whole domain by sampling fewer points, without losing realism in the simulations because the relevant areas are still in high resolution. Another asset of the adaptive sampling is the ability to create volumes without nodes, which correspond to non-karstifiable formations. It is for instance the case of the granite intrusion in the synthetic example used in this paper, through which conduits cannot be simulated (Figure 7). Another advantage of the grid-less approach is that generating random graphs at each iteration enables us to consider the influence of the discretization on the simulation uncertainties (Figure 10). It differs from other methods whose only source of stochasticity comes from a random fracture network generation in the domain.

In relation to the algorithm, the sampling cloud and graph densities are lower in proximity to domain boundaries, potentially leading to less realistic outcomes. A simple workaround consists in defining a sampling domain (and therefore, geological model) slightly larger than the zone of interest. In the examples depicted here, all sampling domains are 5% larger than the probable karstified area. Another asset of using such buffer zone is that the cost function is guaranteed to be well defined at the borders, since geological properties themselves are defined everywhere.

4.2 Number of nearest neighbors

Regarding the numerical accuracy of the simulation with respect to the chosen number of neighbors for the graph, Figure 11 tends to suggest that opting for fewer than 100 neighbors may lead to an average discretization error exceeding 2%. Conversely, employing over 200 neighbors does not enhance simulation realism, as the discretization error stabilizes at approximately 1%. This plateau is likely influenced by the sampling cloud density, which is limited compared to a continuous medium. Therefore, an optimal balance between precision and efficiency for the value of N_{nng} appears to be within the range of 100 and 200 neighbors. This range is approximately equivalent to the number of neighbors in a 3D cubic structured grid, where cells are scanned two levels away in all 26 directions (124 neighbors).

4.3 Vadose/phreatic separation in the karst simulation

In KarstNSim, we add a term to the cost function to separate vadose and phreatic trends, and we separate the vadose and phreatic conduit generation steps. The results (Figure 12) show the interest of explicitly representing vadose conduits in a heterogeneous medium, especially in the presence of an inception horizon. It allows us to avoid imposing an anisotropic cost function to mimic a fictive hydraulic gradient in the vadose zone, as done in 2D in Fandel et al. (2022). Explicitly implementing the inception surfaces is preferable as it increases the

flexibility of the method through the associated cost. As a result, vadose conduits can be influenced by the local gradient of the surfaces topography (strike, dip, folds) at the same time as gravity, creating more realistic patterns.

Explicit water table representation offers more adaptability to the amount of available monitoring well data than an implicit one. If few information is available on the water table position (such as in the Barrois network case study), it can be interpolated from well data with a simple shape, such as a single plane. The advantage of using such explicit surfaces is their flexibility to handle various levels of information.

The water table surface vertically separates vadose and phreatic zones, but is also limited in its x/y extension. Thus, the lateral boundaries of the water table should be carefully defined to avoid aberrant results, especially if the water table surface is not defined explicitly, as in Borghi et al. (2012). KarstNSim handles this complexity by using an explicit 3D triangulated surface for the water table, conformal to the top and base of the aquifer. It is observable in the examples of the Barrois limestones (Figure 19), where the water table associated with the Ribeaucourt spring stops at the contact with the Dommartin and Sublithographic limestones, and does not extend to the Kimmeridgian marls substratum. It has a direct impact, as conduits are not constrained to develop horizontally while they have not reached the base level. However, accurately detecting the exact boundary between vadose and phreatic zones can remain challenging, particularly when surfaces are represented by Triangulated Irregular Networks (TINs), which may occasionally introduce artifacts. These artifacts are evident in the side views presented in Figure 19, where conduits sometimes plunge into the Kimmeridgian marls just before reaching the water table. This issue should be mitigated by increasing the resolution of the triangulated surfaces.

4.4 Implicit fracture representation approach

With an implicit representation of fracturation, the associated cost α_F becomes an easy way to control fracture density. Compared to an explicit approach, the implicit one also saves computing time and memory by skipping the discrete fracture network generation. We choose to impose only 2D orientation trends, but the dip trends could be added to the cost function to create a 3D anisotropy, if dip data are available.

The angular tolerance associated with each fracture family orientation should not be too large, in order to impact the network morphology. Values smaller than 10° permit in practice to produce angular maze patterns.

The implicit approach is however subject to a few restrictions. It is unable to take into account the length of the fractures, which is, of course, finite, whereas conduits can be indefinitely influenced in the same direction in the implicit approach. In practice, however, this effect is often counterbalanced by the regional gradient, which dictates the main conduit orientation. Consequently, conduits will not align with fractures over long distances, unless the fractures are parallel to the hydraulic gradient main orientation.

While our current implicit fracturing approach does not incorporate non-stationary properties, such as variations in fracture orientation or density across the domain, a semi-distributed approach could be implemented to define fracture properties. Following a similar line of thought, distinct weights could be assigned to fractures based on their family, distinguishing between main and secondary orientations, for instance.

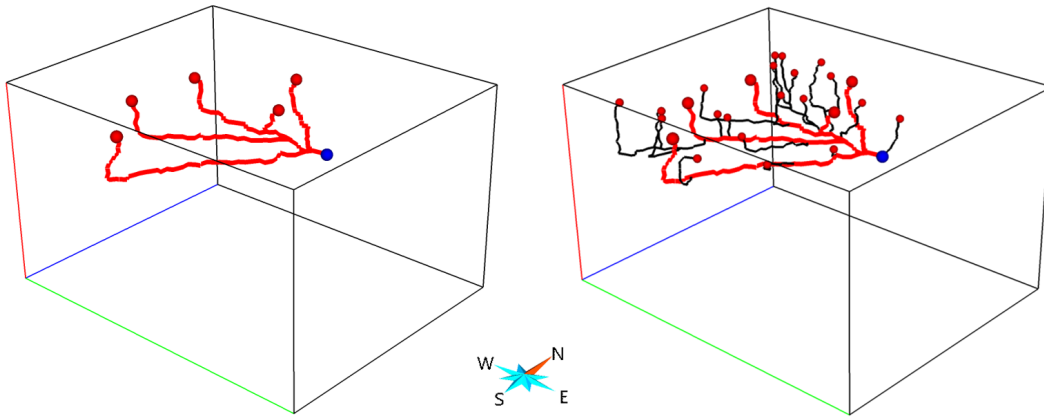


Figure 20 Left: network generated by only taking into account the water table influence with a cost of 1 and $P_{red} = 0.9$. Right: Amplified network (in black), where new inlets have been randomly sampled onto the topography, and connected to the main network (in red) by solving shortest path problems.

4.5 Cohesion of the network

KarstNSim allows user control over the cohesion of the network with the parameter P_{red} (Figure 6). Such idea has been initially proposed in Borghi et al. (2012). In that work, the authors defined a specific, smaller cost for cells crossed by conduits. This smaller cost is not further reduced if conduits crossed the same cells in multiple iterations. To mimic the strong hierarchical organization that can be observed in some branchwork networks, and that is sometimes interpreted as an effect of the positive feedback mechanism of karstification (Dreybrodt and Gabrovšek, 2003), KarstNSim gradually reduces the cost of edges already crossed by conduits. Hence, we can take into account the relative importance of inlets (for example to differentiate sinking rivers to simple dolines) by constraining their iteration order to start with larger, more impactful inlets. Yet, using a randomized ordering and cohesive level in a stochastic frame remains the best choice to assess the effective network uncertainty.

Concerning a range of reasonable values to be picked for P_{red} , Figure 14 shows that values between 0.7 and 0.9 produce hierarchical and cohesive branchwork networks. Yet, those values are based on a five inlet setting. The value of P_{red} should be adapted to the number of inlets, as its incremental application makes its impact proportional to the number of inlets. One interesting perspective would be to operate by groups of inlets.

4.6 Choice of inlets and amplification

The inlet map, defined by aerial photographs or Lidar imaging, naturally strongly impacts the simulation results. Its precision depends on the nature of the landscape (forest, field, road, village...), leading potentially to a biased inlet density.

One way to alleviate this problem consists in selecting only a few inlets, for example by sampling them randomly in a regular grid, as proposed by Malard et al. (2015) and implemented in this study, resulting in the reduction of inlets from 165 to 30 in the simulations of the Ribeaucourt karst network. Several runs of such simulation would help to better assess the network uncertainties.

The networks can then be amplified in one or more other steps (Figure 20) as proposed in the paper of Paris et al. (2021). At each amplification step, new inlets are sampled randomly in the domain, either from existing but still unused inlets, or

from new ones, and conduits are generated by solving shortest path problems between those new inlets and the already existing conduits. This amplification algorithm can also be used to mimic diffuse infiltration when precise inlet points are not well known.

Malard et al. (2015) proposed another way to solve uncertainty on inlet data. They compute the shortest path from the spring to each cell (or node in our case) of the domain, resulting in a cost-distance map. Then, they define the local gradient from each cell, corresponding to the direction to the next cell on the shortest path. Finally, they compute a proxy for the cumulative flow at each cell, corresponding to the number of cells redirecting to it. By skeletonizing the accumulation flow image through a given threshold flow value, they generate a karst network and automatically infer the position of theoretical (but not observed) inlets. This algorithm is similar to the one used to generate theoretical stream channel networks, but it assumes that inlets are placed to optimize today flow drainage in the network, neglecting the fact that karstification occurs throughout multiple steps separated in time, with various hydrogeological contexts. It is also worth noting that the choice of the entry point in their approach is not entirely independent of the model parametrization, as it depends on the meshing. The influence of meshing on surface flow convergence could affect the creation of conduits, particularly when the number of mesh cells drained by a surface point exceeds a specified representation threshold, leading to the initiation of a conduit from that point.

Between inlets, some metrics to choose the iteration order could include their size or their theoretical draining area, determined from present-day Digital Elevation Model (DEM) analysis. Those metrics are, however, limited because they make the assumption that today's hydrogeological setting is a good proxy of the situation during the doline creation, which is not necessarily the case, especially in the cases of polygenic karst systems. For those reasons, it might be best to keep the iteration order between inlets of the same nature random, so that this source of uncertainty can be propagated to the final simulated networks.

4.7 Connectivity uncertainty

The core of the KarstNSim algorithm is based on a shortest path computation between inlets and outlets, to honor the most

common field data we have: sinkhole and spring mapping, and tracer tests. A corollary is that the method is strictly limited to networks where breakthrough (*i.e.*, direct connection between inlets and outlets) has already occurred, and can't be used to dynamically simulate the evolution of cave widening, as would be done in a physically-based method (*e.g.*, Siemers and Dreybrodt, 1998).

In absence of tracer tests, uncertainty can exist on the connectivity between pairs of inlets and outlets of the network, especially close to karst catchment limits. In that case, we proposed to compute the non-Euclidean shortest path distance between the inlet and each possible outlet, and only keep the path leading to the closest outlet. By doing that, we avoid generating unrealistic crossing paths at the catchment limits, which would potentially be obtained if we had kept a randomized connectivity.

4.8 Simulating cave section geometry

After the karst network simulation is achieved, it can be used in a flow simulation model. Yet, in order to do that, the 3D volume of the conduits has to be retrieved. Some authors (Henrion et al., 2008; Collon-Drouaillet et al., 2012; Rongier et al., 2014; Paris et al., 2021) propose different methods to build explicit 3D volumetric envelopes of the conduits. Yet, these approaches are computationally expensive, and are thus best appropriate for small cases or for visualization on the fly. The flow simulators are also not currently adapted for this kind of input. Alternatively, for larger domains of interest, equivalent parameters, such as equivalent radii and width-height ratios, can be used to describe the conduit shapes. Malard et al. (2015) and Borghi et al. (2016) propose to consider the conduit equivalent section parameters to be proportional to their drainage areas. Therefore, they weight inlets accordingly, and the equivalent radius is then propagated from the inlets to the outlet, with constant values along each conduit that increase at each confluence with other conduits. Nevertheless, the hypothesis of a hierarchical repartition of the size of conduits is only acceptable in the case of a complete hydraulic control of the karstogenesis, as opposed to ghost-rock karstification (Dubois et al., 2014). Furthermore, a data analysis on real karst systems made by Frantz (2021) showed that section parameters usually do not follow a particular hierarchical repartition. The curvilinear sequential Gaussian simulation they propose would thus constitute an interesting solution.

4.9 Generation of looping caves

Many pseudo-genetic methods for karst networks generation only reproduce some types of networks, mostly branchwork ones (Collon-Drouaillet et al., 2012; Malard et al., 2015; Luo et al., 2021). Yet, among the four spatial patterns discussed in Jouves et al. (2017), only three of them were reproduced in this paper (vadose branchwork, water table cave, angular passages). The remaining pattern, looping cave, appears in the epiphreatic zone in the case of variations of the water table level (anastomotic patterns), or in the epikarst when recharge is diffuse (angular mazes). Although not extensively tested yet, we are confident in the ability of KarstNSim to generate looping caves, for example through the use of waypoints, as shown by Paris et al. (2021).

5 CONCLUSION

We have proposed a new methodology and code, KarstNSim, to simulate stochastically realistic 3D karst networks in an epigenetic context. With a new cost function, KarstNSim generates a variety of network morphologies driven by the geological knowledge. It avoids most grid artifacts by using a grid-less approach based on an adaptive and efficient random sampling of the 3D space, conformal to geological heterogeneities. Thanks to the randomness of the sampling, the method propagates the discretization uncertainty onto the generated networks in a stochastic workflow. Vadose and phreatic compartmentalization, as well as polygenic karstogenesis are consistently handled.

Different hypotheses on the network morphology can be easily tested, as in the case of the Barrois limestones. The formulation of hypotheses is made simple by the weighted cost function. The networks are simulated consistently with geological data and tracer test observations if available.

Rooms for improvement remain to fully represent the diversity of network patterns, in particular looping caves, dead-end conduits and hypogenic karsts. Ghost-rock karstification could also be considered by modifying the calculation of the average cross-sectional area of the conduits.

CODE AVAILABILITY

KarstNSim is available under MIT license on the following repository : https://github.com/ring-team/KarstNSim_Public. It is provided with a synthetic dataset and a documentation.

ACKNOWLEDGEMENTS

This work was performed in the frame of a research agreement between University of Lorraine, BRGM and ANDRA, and as a part of the "Calcaires du Barrois" program led by ANDRA. The authors would like to thank ANDRA for providing the data used in this study. We would also like to thank for their support the industrial and academic sponsors of the RING-Gocad Consortium managed by ASGA. We acknowledge Emerson for the SKUA-Gocad Software and development kit.

REFERENCES

- Andra (2021). Programme d'activités scientifiques et techniques sur les Calcaires du Barrois Rapport Andra n°CGRPASTR210005. Technical report, Andra. (Cited pages 12, 13, and 15)
- Audra, P and Palmer, A. (2013). The Vertical Dimension of Karst: Controls of Vertical Cave Pattern. In *Treatise on Geomorphology*, volume 6, pages 186–206. Elsevier. (Cited page 10)
- Borghi, A., Renard, P., and Cornaton, F. (2016). Can one identify karst conduit networks geometry and properties from hydraulic and tracer test data? *Advances in Water Resources*, 90:99–115. (Cited pages 1 and 18)
- Borghi, A., Renard, P., and Jenni, S. (2012). A pseudo-genetic stochastic model to generate karstic networks. *Journal of Hydrology*, 414-415:516–529. (Cited pages 2, 5, 10, 16, 17, and 18)
- Bridson, R. (2007). Fast poisson disk sampling in arbitrary dimensions. *ACM SIGGRAPH 2007 Sketches, SIGGRAPH'07*, page 2006. (Cited page 3)
- Caumon, G., Collon-Drouaillet, P., Le Carlier de Veslud, C., Viseur, S., and Sausse, J. (2009). Surface-Based 3D Modeling of Geological Structures. *Mathematical Geosciences*, 41(8):927–945. (Cited page 3)
- Collon, P., Bernasconi, D., Vuilleumier, C., and Renard, P. (2017). Statistical metrics for the characterization of karst network geometry and topology. *Geomorphology*, 283:122–142. (Cited pages 13, 14, and 15)

- Collon, P., Bernasconi, D., Vuilleumier, C., and Renard, P. (2021). Corrigendum to "Statistical metrics for the characterization of karst network geometry and topology". *Geomorphology*, 389. (Cited pages 13 and 15)
- Collon-Drouaillet, P., Henrion, V., and Pellerin, J. (2012). An algorithm for 3D simulation of branchwork karst networks using Horton parameters and A* Application to a synthetic case. *Geological Society Special Publication*, 370(1):295–306. (Cited pages 2, 16, 18, and 19)
- de Rooij, R. (2007). *Towards Improved Numerical Modeling of Karst Aquifers: Coupling Turbulent Conduit Flow and Laminar Matrix Flow under Variably Saturated Conditions*. University of Neuchâtel, Neuchâtel. (Cited page 1)
- Devos, A. (1996). *Hydrologie et aménagements hydrauliques de la Saulx et de l'Ornain, Lorraine*. (Cited page 3)
- Dreybrodt, W. and Gabrovšek, F. (2003). Basic processes and mechanisms governing the evolution of karst. In *Reactions*, pages 1–26. (Cited page 18)
- Dubois, C., Quinif, Y., Baele, J. M., Barriquand, L., Bini, A., Bruxelles, L., Dandurand, G., Havron, C., Kaufmann, O., Lans, B., Maire, R., Martin, J., Rodet, J., Rowberry, M. D., Tognini, P., and Vergari, A. (2014). The process of ghost-rock karstification and its role in the formation of cave systems. (Cited page 19)
- Duran, L. and Gill, L. (2021). Modeling spring flow of an Irish karst catchment using Modflow-USG with CLN. *Journal of Hydrology*, 597(August 2020):125971. (Cited page 1)
- Dwork, N., Baron, C. A., Johnson, E. M., O'Connor, D., Pauly, J. M., and Larson, P. E. (2021). Fast variable density Poisson-disc sample generation with directional variation for compressed sensing in MRI. *Magnetic Resonance Imaging*, 77:186–193. (Cited pages 3 and 7)
- Fandel, C., Miville, F., Ferré, T., Goldscheider, N., and Renard, P. (2022). The stochastic simulation of karst conduit network structure using anisotropic fast marching, and its application to a geologically complex alpine karst system. *Hydrogeology Journal*, 30(3):927–946. (Cited pages 2, 3, 6, 7, and 17)
- Filipponi, M., Jeannin, P. Y., and Tacher, L. (2009). Evidence of inception horizons in karst conduit networks. *Geomorphology*, 106(1-2):86–99. (Cited pages 1 and 9)
- Fournillon, A., Viseur, S., Arfib, B., and Borgomano, J. (2010). Insights of 3D geological modelling in distributed hydrogeological models of Karstic carbonate aquifers. In *Environmental Earth Sciences*, pages 257–262. (Cited page 2)
- Frantz, Y. (2021). *Simulation stochastique des systèmes karstiques*. PhD thesis, Université de Lorraine. (Cited page 19)
- Ghasemizadeh, R., Hellweger, F., Butscher, C., Padilla, I., Vesper, D., Field, M., and Alshawabkeh, A. (2012). Review: Groundwater flow and transport modeling of karst aquifers, with particular reference to the North Coast Limestone aquifer system of Puerto Rico. *Hydrogeology Journal*, 20(8):1441–1461. (Cited page 1)
- Hart, P. E., Nilsson, N. J., and Raphael, B. (1968). A Formal Basis for the Heuristic Determination of Minimum Cost Paths. *IEEE Transactions on Systems Science and Cybernetics*, 4(2):100–107. (Cited page 2)
- Henrion, V., Pellerin, J., and Caumon, G. (2008). A stochastic methodology for 3D cave systems modeling. In *Geostats 2008*, pages 525–533, Santiago, Chile. Eighth Geostatistical Congress. (Cited pages 2 and 18)
- Jaillet, S. (1995). *Le géosystème karstique du Rupt du Puits. étude morphologique de l'exo et de l'endokarst*. PhD thesis, Université de Metz. (Cited page 3)
- Jaquet, O., Siegel, P., Klubertanz, G., and Benabderrhamane, H. (2004). Stochastic discrete model of karstic networks. *Advances in Water Resources*, 27(7):751–760. (Cited pages 2 and 3)
- Jouves, J., Viseur, S., Arfib, B., Baudement, C., Camus, H., Collon, P., and Guglielmi, Y. (2017). Speleogenesis, geometry, and topology of caves: A quantitative study of 3D karst conduits. *Geomorphology*, 298:86–106. (Cited pages 2 and 19)
- Kanfar, R. and Mukerji, T. (2023). Stochastic Geomodeling of Karst Morphology by Dynamic Graph Dissolution. *Mathematical Geosciences*. (Cited page 13)
- Kiraly, L. (1988). Large Scale 3-D Groundwater Flow Modelling in Highly Heterogeneous Geologic Medium. In *Groundwater Flow and Quality Modelling*, pages 761–775. Springer Netherlands, Dordrecht. (Cited page 1)
- Luo, L., Liang, X., Ma, B., and Zhou, H. (2021). A karst networks generation model based on the anisotropic Fast Marching algorithm. *Journal of Hydrology*, 600:126507. (Cited pages 2 and 19)
- Malard, A., Jeannin, P. Y., Vouillamoz, J., and Weber, E. (2015). Approche intégrée pour la délimitation des bassins d'alimentation et la modélisation du réseau de conduits des systèmes karstiques: application au jura tabulaire Suisse. *Hydrogeology Journal*, 23(7):1341–1357. (Cited pages 18 and 19)
- Mangin, A. (1974). Contribution à l'étude hydrodynamique des aquifères karstiques. Technical report, Université Dijon. (Cited page 1)
- Mazzilli, N., Guinot, V., Jourde, H., Lecoq, N., Labat, D., Arfib, B., Baudement, C., Danquigny, C., Dal Soglio, L., and Bertin, D. (2019). KarstMod: A modelling platform for rainfall - discharge analysis and modelling dedicated to karst systems. *Environmental Modelling & Software*, 122:103927. (Cited page 1)
- Palmer, A. (1991). Origin and morphology of limestone caves. *Geological Society of America Bulletin*, 103(1):1–21. (Cited page 1)
- Palmer, A. (2003). Speleogenesis in carbonate rocks. *Speleogenesis and Evolution of Karst Aquifers*, 1(1):1–11. (Cited page 5)
- Paris, A., Guérin, E., Peytavie, A., Collon, P., and Galin, E. (2021). Synthesizing Geologically Coherent Cave Networks. *Computer Graphics Forum*, 40(7):277–287. (Cited pages 2, 3, 5, 13, 15, 16, 18, and 19)
- Rongier, G., Collon-Drouaillet, P., and Filipponi, M. (2014). Simulation of 3D karst conduits with an object-distance based method integrating geological knowledge. *Geomorphology*, 217:152–164. (Cited page 18)
- Sethian, J. A. (1996). A fast marching level set method for monotonically advancing fronts. *Proceedings of the National Academy of Sciences of the United States of America*, 93(4):1591–1595. (Cited page 2)
- Sethian, J. A. (2001). Evolution, Implementation, and Application of Level Set and Fast Marching Methods for Advancing Fronts. *Journal of Computational Physics*, 169(2):503–555. (Cited page 2)
- Siemers, J. and Dreybrodt, W. (1998). Early development of Karst aquifers on percolation networks of fractures in limestone. *Water Resources Research*, 34(3):409–419. (Cited pages 2 and 18)
- Sivelle, V., Renard, P., and Labat, D. (2020). Coupling SKS and SWMM to solve the inverse problem based on artificial tracer tests in karstic aquifers. *Water (Switzerland)*, 12(4). (Cited page 1)
- Thiéry, D. (1982). Improvement of methods of long term prediction of variations in groundwater resources and regimes due to human activity. *Hydrological Sciences Journal*, 27(2):216–229. (Cited page 1)
- Wyvill, B., Guy, A., and Galin, E. (1999). Extending the CSG Tree. Warping, Blending and Boolean Operations in an Implicit Surface Modeling System. *Computer Graphics Forum*, 18(2):149–158. (Cited page 4)