Mathematical Geology, Vol. 38, No. 5, July 2006 (© 2006)
DOI: 10.1007/s11004-006-9033-0

Modelling Spatial Variability Along Drainage Networks with Geostatistics[1]

Jean-Stéphane Bailly,[2] Pascal Monestiez,[3] and Philippe Lagacherie[4]

Local characteristics of drainage networks such as cross-section geometry and hydraulic roughness coefficient, influence surface water transfers and must be taken into account when assessing the impact of human activities on hydrological risks. However, as these characteristics have not been available till now through remote sensing or hydrological modelling, the only available methods are interpolation or simulation based on scarce data. In this paper we propose a statistical model based on geostatistics that allows us to take account of both the spatial distribution and spatial uncertainties. To do this, we modify the geostatistical framework to suit directed tree supports corresponding to drainage network structures. The stationarity concept is specified assuming conditional independence between parts of the network; variogram fitting and modelling are then modified accordingly. A sequential multi Gaussian simulation procedure going upstream along the network is proposed. We illustrate this approach by studying the width of an 11-km long artificial drainage network in the south of France.

KEY WORDS: directed tree, geostatistical modelling, upstream-downstream stationarity, simulation, cross-section geometry.

INTRODUCTION

Drainage networks are sets of connected linear features (e.g. field drainage pipes, ditches, channels, storm water pipes, rivers) in which the water flows towards the outlet of a given catchment. They are seen by hydrologists as links between areas producing water, pollutants, sediments and water resources or freshwater systems such as groundwater tables, rivers or lakes. So it is important to take account of these drainage networks when assessing all the possible impacts of

[1]Received 15 Septmber 2003; accepted 7 June 2005; Published online: 26 September 2006.
[2]UMR TETIS, ENGREF-Cemagref-CIRAD, 500 rue Jean-François Breton, 34093 Montpellier Cedex 5, France; e-mail: bailly@teledetection.fr.
[3]Unité de Biométrie, INRA, Domaine St Paul, Site Agroparc, 84914 Avignon Cedex 9, France; e-mail: monestiez@avignon.inra.fr.
[4]UMR LISAH, 2 place Viala, INRA, 34060 Montpellier Cedex 01, France; e-mail: lagache@ensam.inra.fr.

0882-8121/06/0700-0515/1 © 2006 International Association for Mathematical Geology

human activities on hydrological risks, e.g. flooding, surface water pollution, eutrophication and silting of freshwater systems.

For most hydrologists, spatially distributed modelling is the best tool for assessing the impact of these human activities (Beven, 1996, p. 293) while "we want to know not only what is the quality and quantity of water in a stream, but also where any contaminants came from and where best to invest scarce financial resources to help rectify the problem" (Grayson and others, 2002, p. 1314). In order to identify the factors that have the most impact and to study the consequences of local decisions (e.g. changes of an agricultural practice, building a road, rivers excavation, etc.) on the catchment hydrology we need to know how much variability there is within the catchment due to topography, soil, anthropogenic features and hydrological variables. Some recent spatial models now include the water transfer through drainage networks using physical derivative forms of the Saint-Venant equations (Moussa and others, 2002; Varado, 2004, p. 115; Estupina-Borrell, Llovel and Dartus, 2002; Carluer and de Marsily, 2005).

However, the lack of data describing the spatial variability within the catchment often limits the application of spatial models (Beven, 2001, p. 5; Beven and Feven, 2002; Grayson and others, 2002, p. 1315). This is particularly true for the variability of local drainage network characteristics, such as cross-section geometry or roughness coefficient, that govern water transfers. As remote sensing technology does not provide accurate estimates of these characteristics, they are often obtained by fitting parameters to hydrological models (Moussa and others, 2002). Leopold and Maddock (1953) proposed a simple model of spatial variability in which the cross-section geometry (width, height) increases linearly on a logarithm scale, with the size of the contributing area. However, the few studies that actually describe real drainage networks (Lagacherie and others, 2005; Bouldin and others, 2004; Raff, Smith and Trlica, 2003) showed much more complex patterns of variations. So it is vital to develop new methods that could provide more realistic models of the variability of drainage networks that could be integrated into spatially distributed models.

In a recent review, Grayson and others (2002) explored several strategies for describing the spatial variability of distributed model parameters and hydrological variables. Stochastic simulation appears to be a good way of proceeding when data is scarce. For example, two studies have used stochastic simulation of parameters in space to study the sensitivities of hydrological model respectively to spatial distribution (Grayson and others, 2002, p. 1317) and spatial uncertainties (Chilès and Delfiner, 1999, p. 453). One drawback of these methods is that they require fully explicit probabilistic models fitted to the available data. Up to now these methods have not been applied on drainage networks (Western, Blöschl and Grayson, 2001).

In this paper we propose an approach for simulating the spatial distribution of local characteristics along a drainage network. This approach must deal with

the specificity of the network considered as a directed tree and must model characteristics not only on the network nodes but also at any site along the linear section between nodes. Unlike Markovian (Guedon, 2001) or lattice models (Cressie, 1993, p. 383) that focus on what happens only at nodes, geostatistics which deals with random functions on a continuous support seems to be a more suitable theoretical framework. Another advantage is that some methods for spatial fields simulation already exist (Lantuéjoul, 2002, p. 189). However, as geostatistics was initially designed for domains in two or three dimensional spaces, it must be adapted to fit the directed tree support. Some geostatistical approaches on trees have already been developed but not on directed ones. For example, Audergon, Monestiez and Habib (1993) and Monestiez, Habib and Audergon (1989) have applied geostatistics to fruit trees to characterize the spatial variability of the size of the fruit. Introducing a direction along the network requires a different set of assumptions. So we start by reviewing the basics: firstly the stationarity hypothesis and secondly, the modelling and fitting of the spatial structure through a variogram or a covariance function.

In this paper, we first introduce the theoretical framework required to specify geostatistical hypotheses and models on directed trees. We then propose a procedure for simulating random variables on directed trees. The case study selected to illustrate our approach deals with the variability of the ditches width along an 11-km long artificial drainage network located in a small cultivated catchment area (1-km^2) in the south of France (Louchart and others, 2001).

MODELS AND METHODS

Definitions and Notations on Drainage Network

To introduce the usual hypotheses and concepts of geostatistics on drainage networks, we start by few definitions and some notations. We have to define two kinds of directed tree models to represent the drainage network under study. The first one uses standard concepts from graph theory and is used to characterize topological relationships between elementary units of the network. The second is seen as a geometrical support for random functions; it consists of line segments that are linked together to build the topology in question.

Vertices and Edges

The drainage network is first modelled by a *directed and valued tree* G.
$G = [X, U]$ is a tree, i.e. a connected graph without circuits, with vertices i and edges u. Let X denote the set of vertices and U the set of edges.

Every edge u corresponds to a river segment linking two vertices and is naturally directed by water flow from upstream to downstream. We denote an

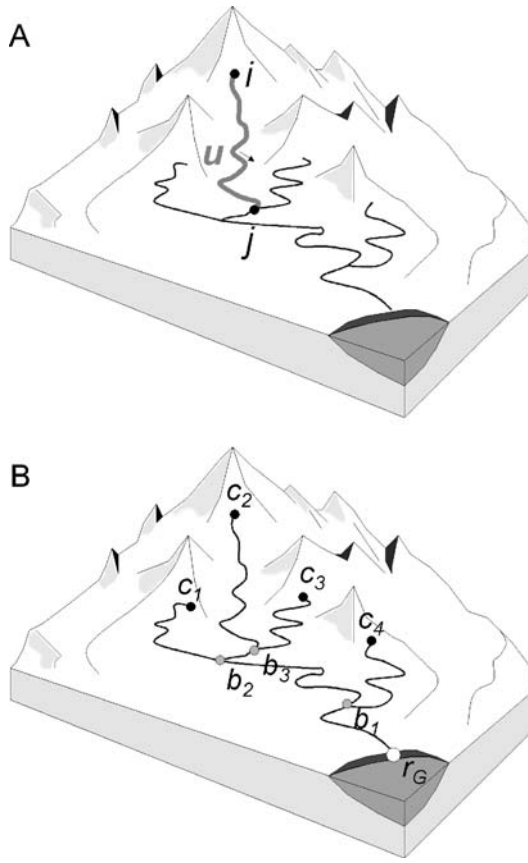

Figure 1. Example of drainage network. (A) a directed edge $u = (i, j)$. (B) the set of sources c_1, c_2, c_3, c_4 (black), the set of junctions b_1, b_2, b_3 (gray) and the outlet r_G (white).

edge $u = (i, j)$, where i, j are vertices and i is u's initial extremity, corresponding to the upstream vertex and j is u's final extremity, corresponding to the downstream vertex (Fig. 1A).

For each vertex i, let $d_G(i)$ be the *degree*, equal to the number of distinct edges having i as its extremity. A vertex of degree one, i.e. an extremity of the directed tree, is called *source* when it is an initial extremity of an edge, and *outlet* when it is a final extremity of an edge (Fig. 1B).

From these definitions, it follows that there can be only one outlet. In what follows this vertex is chosen to be the *root* of the directed tree and denoted r_G (Fig. 1B).

Vertices with degree d_G strictly greater than 2 are called *junctions* (Fig. 1B).

The set of sources is denoted S and the set of junctions B. Let N_S be the number of sources.

Directed Paths and Rivers

A directed path between two vertices (k, l) corresponds to the water course on the drainage network from k, the upstream vertex to l, the downstream vertex through successive edges (Fig. 2A). So directed paths are defined only for vertices that are in an upstream-downstream relationship, i.e. along the same hydrographic axis. Let (k, l) be the path L between k and l. This generalizes the edge notation while L is an edge sequence. We expand on the definition of a directed path in appendix A.

Edge v is a downstream edge of u if there exists a path containing u and v with u before v. Let Γ_u be the set of all edges that are downstream edges of u. Similarly, v is a upstream edge of u if u belongs to Γ_v. Γ_u^{-1} is the set of all upstream edges of u.

We define a *river* as a directed path that connects a source to a junction. It is always possible to decompose the whole drainage network in a set of rivers A_o (Fig. 2B) that is a partition of the network, i.e. each edge belongs to a river and only one. The number of rivers in A_o equals the number of sources N_S.

Edge Valuation, Lengths and Ordering

We can also consider G as a *valued* graph, by associating a value (its length measured on the actual drainage network), to each edge.

For all $u \in U$, the edge length is denoted $l(u)$. By extension, a length l_{ij} is defined for any path (i, j) by summation:

$$\forall L = (i, j) = \{u_1, u_2, \ldots, u_q\} \quad l_{ij} = l(L) = \sum_{k=1}^{q} l(u_k).$$

Consequently an ordering on the tree vertices X is defined by the lengths of paths between any vertex and the outlet r_G. We can then deduce a river order from the order on junctions because there is a one-to-one mapping between junctions and rivers (appendix A.2). An example of junctions order on B, $\{b_1, b_2, b_3\}$ is given in Fig. 1B. The A_o set becomes an ordered sequence of rivers:

$$A_o = \left(a_{(1)}, a_{(2)}, \ldots, a_{(N_S)}\right), \tag{1}$$

where the first one $a_{(1)}$ is associated with the outlet and the last one with the junction that is the farthest from the outlet. A more detailed example of how to create an ordered series of rivers is given in appendix A.

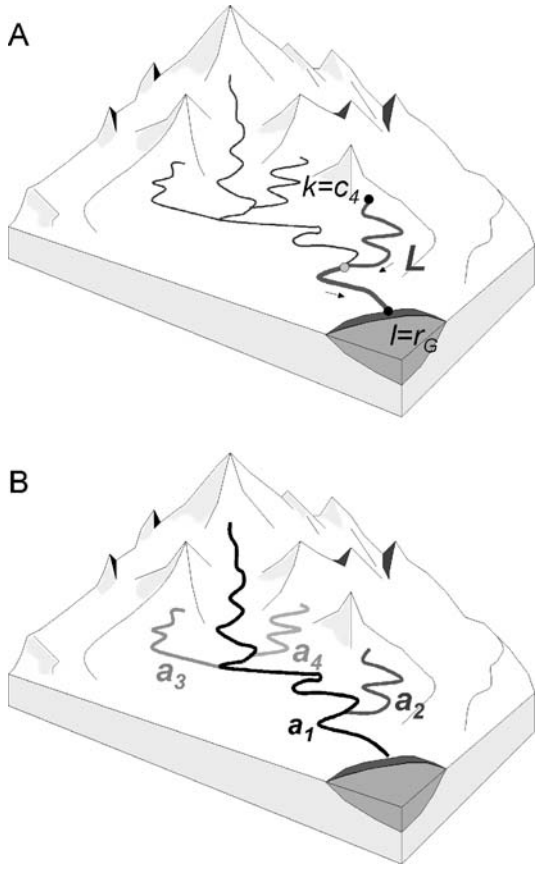

Figure 2. Example of directed tree. (A) a directed path $L = (k, l) = (c_4, r_G)$. (B) a set of rivers $A_o = \{a_1, a_2, a_3, a_4\}$.

The Directed Tree as a Continuous Support of the Spatial Process

We now consider the edges as line segments $[i, j]$ of the drainage network, i.e., as a geometric set of points x, together with a distance d which is the curvilinear distance along the rivers axes. To avoid confusion with G, this tree is denoted G_M. By extension, the vertices i, j are also points in G_M and $d(i, j) = l_{i,j}$.

For two points (x, y) in G_M (Fig. 3A), with $x \in u = (i, j)$, $y \in v = (k, m)$ and with $u \in \Gamma_v^{-1}$, the distance between x and y is defined by

$$d(xy) = d(xj) + l_{jk} + d(ky).$$

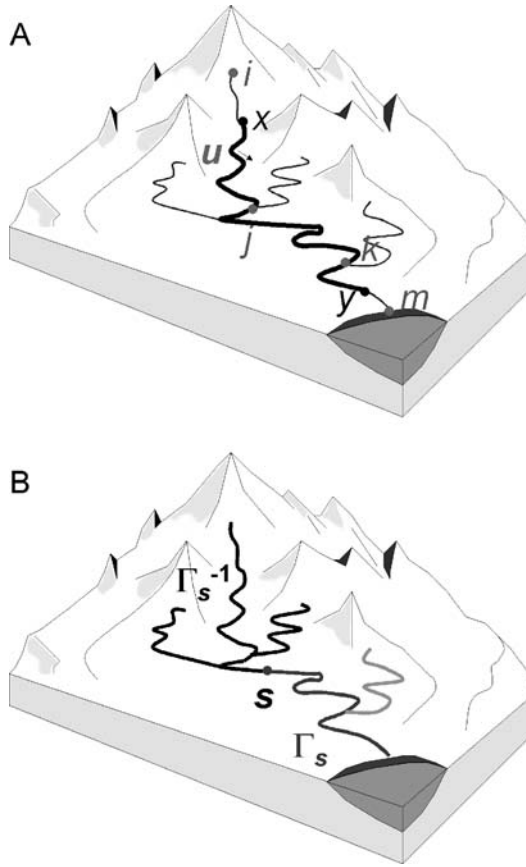

Figure 3. Example of directed tree. (A) Distance definition on G_M: distance $d(xy)$ between two points x and y that belong to a same water course. (B) illustration of sets Γ_s, set of downstream points (dark grey), Γ_s^{-1}, set of upstream points (black), and the complement on G_M to Γ_s^{-1} and Γ_s (light grey), the set of "parallel rivers" for a point s.

Conversely, if $u \in \Gamma_v$, we get $d(xy) = d(xi) + l_{im} + d(my)$.

If $u \notin \Gamma_v \bigcup \Gamma_v^{-1}$, then $d(x, y)$ is undefined; we say that x and y lie on "parallel" rivers.

For all x in G_M, we generalize the notion of upstream edges to upstream points (Fig. 3B) : $x \in u$ with $u = (i, j) \in U$,

$$\Gamma_x^{-1} = \{y \mid y \in [i, x]\} \cup \{y \mid y \in v \text{ and } v \in \Gamma_u^{-1}\}$$

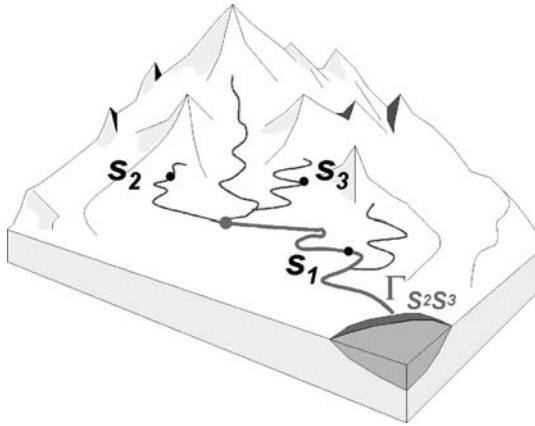

Figure 4. Stationarity on G_M: Example of points s_2, s_3 at equal distance from a same downstream point s_1 and $\Gamma_{s_2 s3}$ in light grey, set of common downstream points of s_2, s_3.

Conversely, $\Gamma_x = \{x\} \cup \{y \mid x \in \Gamma_y^{-1}\}$ defines the set of downstream points of x.

Further, only G_M which is the geometrical support for stochastic processes on the drainage network under study is used.

Geostatistical Models on a Drainage Network

Classical geostatistical models are based on strict stationarity, second-order stationarity or the intrinsic hypothesis.

Let us define $Z(s)$ as a spatial random function on G_M. Strict stationarity, i.e. invariance of the spatial distribution under translation, is defined for all integers n by
$\forall z_1, \ldots, z_n$, $\forall s_1, \ldots, s_n$, and $\forall s'_1, \ldots, s'_n$ such as $s'_1 \in \Gamma_{s_1}, \ldots, s'_n \in \Gamma_{s_n}$ and such as $d(s_1, s'_1) = \ldots = d(s_n, s'_n) = h$, the two probabilities are equal :

$$P(Z(s_1) < z_1, \ldots, Z(s_n) < z_n) = P(Z(s'_1) < z_1, \ldots, Z(s'_n) < z_n), \qquad (2)$$

that corresponds to invariance of the distribution under upstream-downstream translation of all n points.

Equation (2) leads to identical random variables $z(s)$ at any points that are equidistant from the outlet. For example, in Fig. 4, s_2 and s_3 are equidistant points from s_1 then they correspond by downstream translation to s_1. If we apply equation

(2) to this example, it becomes

$$\forall z_2, z_3, \quad P(Z(s_2) < z_2 \ , \ Z(s_3) < z_3) = P(Z(s_1) < z_2 \ , \ Z(s_1) < z_3).$$

Consequently, random function $Z(s)$ is identical for s_2 and s_3. This may be reproduced upstream or downstream if there is a known downstream point so knowing the random function on the main river (which is supposed to be the longest one) effectively determines the random function on the whole tree.

Consequently stationarity cannot be extended in a straightforward manner to directed tree-shaped supports. We have to introduce an additional hypothesis related to the tree topology and the tree direction.

Spatial Distribution, Conditional Independence and Stationarity

This additional hypothesis on geostatistical models for stream networks is called *conditional independence between parallel rivers* and is explained in appendix B.

This hypothesis is illustrated by the example in Fig. 4. Let $\Gamma_{s_2 s_3}$ be the intersection of Γ_{s_2} and Γ_{s_3}, i.e. the set of downstream points common to s_2 and s_3.

If $Z(\Gamma_{s_2 s_3})$ denotes the random function Z for all points that belong to $\Gamma_{s_2 s_3}$, then the two conditional distributions are equal:

$$\mathcal{L}(Z(s_2)|Z(s_3), Z(\Gamma_{s_2 s_3})) = \mathcal{L}(Z(s_2)|Z(\Gamma_{s_2 s_3})) \tag{3}$$

Conditional independence between parallel rivers means that random variables on parallel rivers are independent once the random function in the common downstream part of these parallel rivers is known. This corresponds to a downstream causality hypothesis. It comes straight from the direction of the tree structure and is highly suitable for variables that are linked to water flow (e.g., a pollution rate in a river network).

The stationarity is then simply deduced for any directed path (appendix B) and corresponds to invariance by translation upstream-downstream, i.e. invariance by translation on water courses.

Weak Stationarity

When the previous hypothesis is limited to only the first two moments, we obtain

$$\forall x \in G_M, \mathrm{E}[Z(x)] = m, \tag{4}$$

where E denotes the mathematical expectation and,
$\forall s, s' \in G_M^2$ such as $s' \in \Gamma_s$ and $d(s, s') = h$,

$$\mathsf{Cov}(Z(s), Z(s')) = \mathsf{E}[Z(s)Z(s')] - m^2 = C(h). \tag{5}$$

A covariance function $C(h)$ is a positive definite function in one dimension space.

The covariance $\mathsf{Cov}(Z(s), Z(s'))$ when $s' \notin \Gamma_s \cup \Gamma_s^{-1}$ is not zero and cannot be expressed by a simple function of the relative locations of s and s' on the tree. Only the conditional covariance is zero.
If $\Gamma_{ss'} = \{s''|s'' \in \Gamma_s \cap \Gamma_{s'}\}$, then

$$\mathsf{Cov}(Z(s), Z(s')|Z(s''), s'' \in \Gamma_{ss'}) = 0$$

Similarly, we define the variogram:

$$\forall s, s' \in G_M^2 \quad \text{such as} \quad s' \in \Gamma_s \quad \text{and} \quad d(s, s') = h,$$
$$\gamma(h) = \frac{1}{2}\mathsf{E}[(Z(s) - Z(s'))^2]$$

Drift Models

For more general processes, let $m(s)$ denote a drift. It corresponds to a non-stationary expectation of $Z(s)$. The model $Z(s)$ becomes

$$\forall s_\alpha, \quad Z(s_\alpha) = m(s_\alpha) + Y(s_\alpha) \tag{6}$$

where $m(s)$ is a deterministic function of $x(s)$, $x(s)$ is a location index on the tree structure and $Y(s)$ a centered stationary random function.

A location index $x(s)$ represents the s location in the tree. Several possible location indices exist. For example, they may be:

- $x_1(s) = d(s, r_G)$ the distance to the root,
- $x_2(s) = \text{argmax}_{c_l \in S} d_{c_l,s}$ the maximum distance to a source,
- $x_3(s) = \sum_{v \in \Gamma_u} l(v) + d(i, s)$ the upstream cumulative linear distance, where $u = (i, j)$ is the edge to which s belongs.

The final choice of a location index to model the drift depends on the underlying physical processes that govern the long range variations, and these differ from one variable to another.

Experimental Variogram

If n locations have been sampled and $Z(s_1), \ldots, Z(s_n)$ measured, the definition of the experimental variogram is generalized on G_M by:

$$\gamma^*(h) = \frac{1}{2\,N_h} \sum_{s_\alpha - s_\beta \simeq h} (z(s_\alpha) - z(s_\beta))^2 \tag{7}$$

where N_h denotes the count of pairs (s_α, s_β) separated approximately by the lag h and such as s_β belongs to Γ_{s_α}.

Simulating a Random Function on a Drainage Network

The random function $Z(s)$ on the drainage network G_M is simulated for a finite set $s_k, k = 1, \ldots, N$ of points which is dense enough to represent spatial variations all over the network. We first remove the drift model or the constant mean to work on the simulation of $Y(s_k)$ that is supposed to be normally distributed and centered. The classical procedure, direct simulation based on the covariance matrix between points s_k (Chilès and Delfiner, 1999, p. 465), cannot be used here. However, it is possible to use a sequential simulation procedure based on the conditional independence and the upstream-downstream covariance function.

We started from the decomposition in rivers $A_o = (a_{(1)}, a_{(2)}, \ldots, a_{(N_S)})$ previously defined in equation (1).

We first simulate points on the main river $a_{(1)} = (c_1, r_G)$. The covariance matrix is deduced from $C(h)$. All methods that can be simulated $Z(s)$ for a reasonable number of points may be used here. We used the approach based on the Cholesky decomposition of the covariance matrix (Ripley, 1987, p.98).

To simulate $Z(s)$ for points belonging to $a_{(2)} = (c_2, b_{(2)})$ – where c_2 and $b_{(2)}$ are respectively a source and a junction – we condition their simulation to points that belongs to $\Gamma_{b_2} \cap a_{(1)}$, i.e., points downstream the junction $b_{(2)}$. They all belongs to a single directed path (c_2, r_G). The covariance function is then defined for all pairs and any method for conditional simulation may be used. We used the one which uses kriging for conditioning to already simulated data initially described by Journel and Huijbregts in Cressie (1993), p. 207.

We then iterate the procedure following the sequence of rivers in A_o (Fig. 5). At each step we condition the values of points on river $a_{(i)}$ by the values of points that have already been simulated and that belong to the downstream path $\Gamma_{b_{(i)}}$ of the corresponding junction $b_{(i)}$ of $a_{(i)}$.

The simulation of all points is completed when points on the last river $a_{(N_S)}$ have been simulated. A more detailed description of this simulation procedure is presented in appendix C.

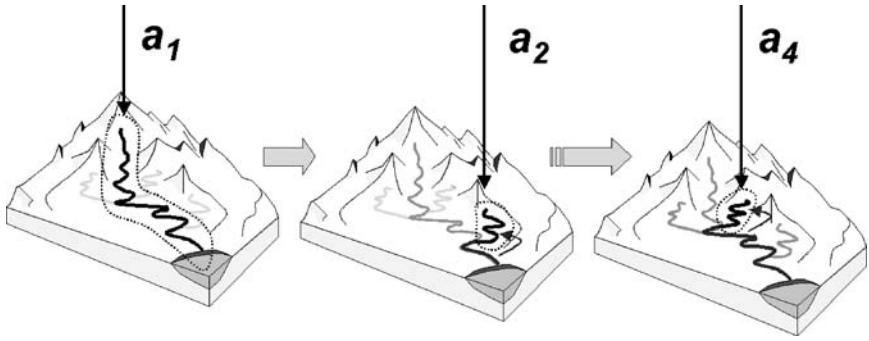

Figure 5. Simulation of random function on drainage network: procedure is iterated for each river simulation from downstream to upstream: *step 1* – simulation of main river a_1 (black) ; *step 2* – simulation of river a_2 (black) conditionally to points simulated on a_1 downstream a_2 (dark grey); *last step* - simulation of the upper river a_4 (black) conditionally to all points simulated on segments of rivers downstream a_4 (dark grey).

CASE STUDY

We have applied geostatistical models on directed trees to an artificial drainage network composed mainly of ditches in a 1-km^2 cultivated catchment, the Roujan catchment. It is part of the Hérault river basin in the south of France. This catchment is intensively cultivated, mainly vineyards, and is structured in three different morphological units: a small upstream area on a plateau, a steeply sloping middle section with artificial terraces and a gently sloping downstream area (Louchart and others, 2001) as seen on Fig. 6A.

Data

The 11 km of the artificial drainage network of the Roujan catchment have been exhaustively studied. The length of ditches and their connections have been described and stored in a geographical database (Lagacherie and others, 2005). The geometry and directed topology of the Roujan network defines a graph G, which contains 62 sources, 52 junctions and 1 outlet.

We selected a set of sites on the network. This selection, which included junctions, resulted from choices based on edge orientation and length. Each change in section orientation delineated a site. We decided on a maximum length of 10 meter for straight sections between sites. $\mathcal{D} = \{s_1, \ldots, s_\alpha, \ldots, s_N\}$ is the set of points on G_M resulting from this selection of sites. It contains 1609 vertices, 1608 edges, 62 sources and 52 junctions.

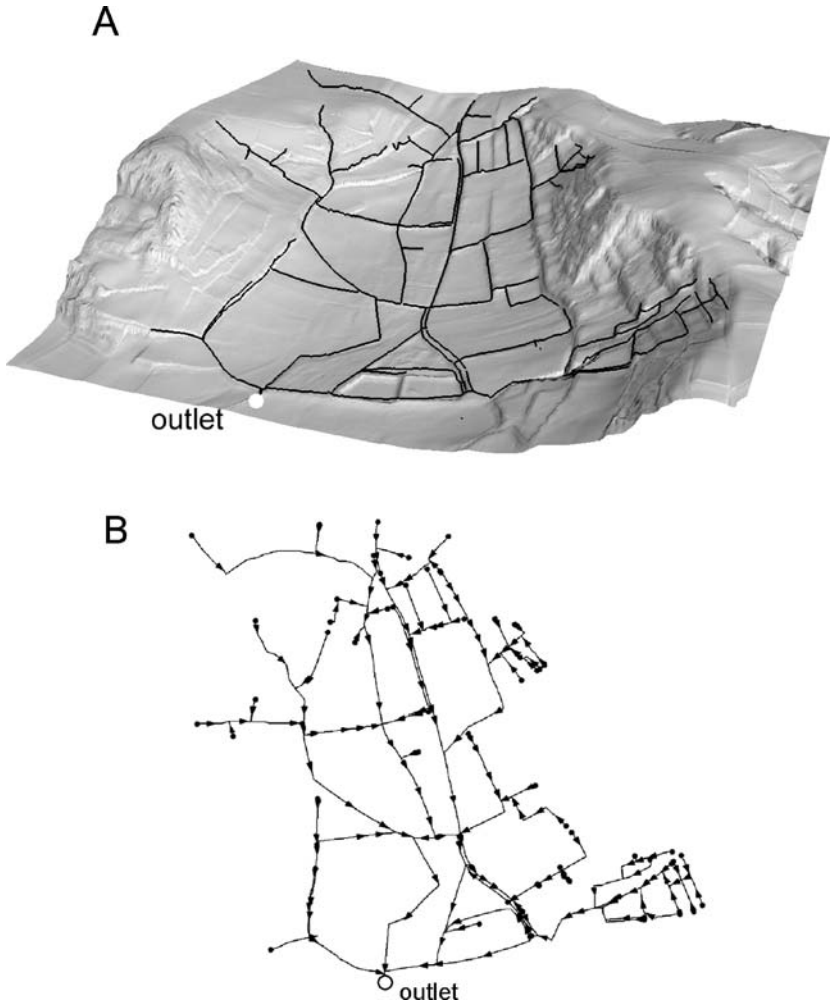

Figure 6. The artificial drainage network of the Roujan catchment (area of about 1 km^2). (A) depicted in 3 dimensions on a Digital Elevation Model with a vertical scale multiplied by 5. (B) in geographic space (the arrows represent the upstream-to-downstream direction of each edge).

We excluded 173 sites, representing 11 % of $\mathcal{D}$'s cardinal, for future analysis. They correspond to particular sections on the artificial drainage network:

– roads: even if they link parts of the network, roads are not meant for water transport and their properties are not significant for the analysis.

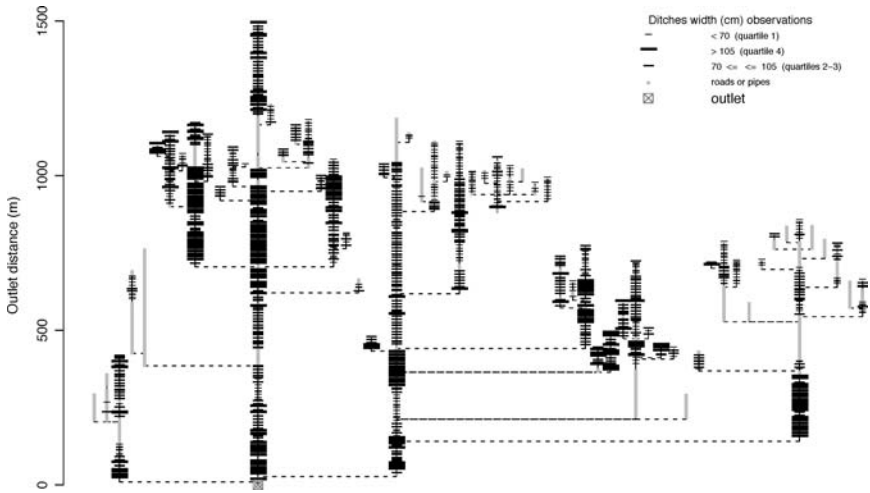

Figure 7. Width data depicted by quartiles on a topological scheme of the Roujan artificial drainage network : each sample is depicted by a dash with an abscissae given by its position from right to left in the branching structure and with ordinate corresponding to its distance from the root.

– pipes: their properties remain constant; we do not need to model and simulatethe spatial distribution of their properties.

Qualitative and quantitative variables were measured at 1136 sampled sites, s_α, which correspond to a 79% sampling rate. We present results for only one hydraulic variable: the ditch width, in particular, the mean of the widths at the top and bottom of the ditch. We chose this variable because of its measurement stability and because it is representative of hydraulic dimensional parameters for ditches. Note that the widths at the top and bottom of ditches were measured by the same observer with a precision of about 10 cm.

Description of the Data Set

Figure 6B presents these data as a graph in geographic space. As it may be difficult to understand the topology and geometry structures of networks in this way, we also present it as a G tree representation (Fig. 7). The x coordinate of each vertex s_α, corresponds to a river number that respects left and right position on junctions regarding direction while the y coordinate corresponds to $x_1(s_\alpha)$ defined above, the distance to the root.

Figure 7 does not show any particular spatial distribution of ditch widths, except for some weak values near the sources. It also does not show any significant increase in the drift from upstream to downstream.

DATA MODELLING AND SIMULATION

A Test for the Absence of Spatial Structure

To test the significance of the spatial structure in the data we first have to specify the null hypothesis **H0** which corresponds to complete spatial randomness, i.e. the absence of spatial structure:

$$\mathbf{H0}:$$
$$\mathsf{E}(Z(s_\alpha)) = m(s_\alpha) = m, \quad \forall s_\alpha \in G_M$$
$$\text{and independent } (Z(s_\alpha), Z(s_\beta)), \forall (s_\alpha, s_\beta) \in {G_M}^2$$

A Mantel test (Legendre and Legendre, 1998, p. 553) was used. We computed the standardized Mantel statistic between the inverse distance matrix (with distances presented above) and a dissimilarity (square difference) matrix between sampled locations on the network. One hundred thousand permutations were computed to simulate the distribution of the test statistic under **H0** for a risk level $\alpha \leq 0.05$ (Manly, 1997, p. 174). The standardized Mantel statistic (-0.04) had a probability ($P < 0.0001$) that significantly rejects **H0**, i.e. absence of spatial structure.

So we can now estimate and model an upstream-downstream drift structure $\mathsf{E}(Z(s_\alpha)) = m(s_\alpha) = f_\theta(x(s_\alpha)), \quad \forall s_\alpha \in G_M$.

Drift Modelling

Location Index for Upstream-to-downstream Drift

The analysis of $m(s_\alpha)$ is carried out in the upstream-downstream direction. To be able to locate each site where ditch widths have been observed in this direction, we define an upstream-downstream location index $x(s_\alpha)$ for each vertex s_α in a graph [Eq. (8)], as presented above, corresponds to x_3, the total curvilinear length of edges up to s_α.

$$x_3(s) = \sum_{v \in \Gamma_u} l(v) + d(i, s) \tag{8}$$

This location index has been selected as the ditch width variable. Its main advantage is that it takes account of different sources that are upstream of a site. So it cumulatively integrates the surface of the drainage area from upstream to downstream. This is an explicative hydraulic parameter underlying the width

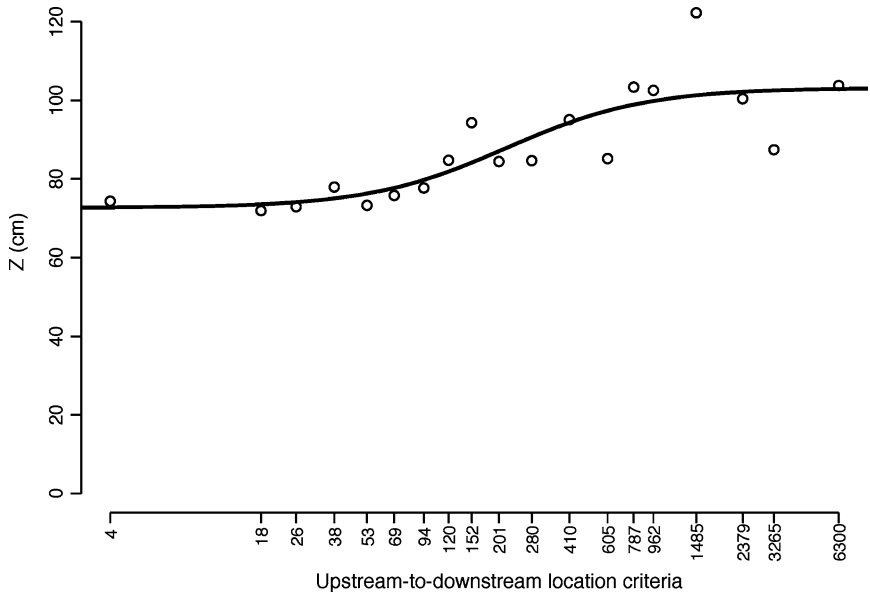

Figure 8. Experimental (points) and fitted (line) for upstream-to-downstream drift.

variable. For other variables, other criteria such as x_1 or x_2 (see definitions above) may be more suitable to represent an upstream-to-downstream drift.

We first grouped s_α sites into classes based on $x_3(s_\alpha)$ upstream-to-downstream location index. Class breaks were defined by quantiles on the number of vertices, except for the first class which includes all sources and sites within 15 m from sources. For each class, the empirical mean was calculated on the width data.

Figure 8 shows on a logarithmic scale for $x_3(s_\alpha)$ $\mathsf{E}(Z(s_\alpha)) = m(s_\alpha) = m$, $\forall s_\alpha \in G_M$, with empirical drift depicted by points.

Figure 8 shows two sills with a change between them. We fitted an inverse logistic model for drift on observed data on a logarithmic scale while respecting sills and the change:

$$m(s_\alpha) = \theta_1 \frac{e^{(\theta_2(p(s_\alpha)-\theta_3))}}{1 + e^{(\theta_2(p(s_\alpha)-\theta_3))}} + \theta_4$$

with $p(s_\alpha) = log(x_3(s_\alpha))$.

The weighted least-square estimate of θ (Fig. 8), with weights equal to the number of data points in each class of the upstream-downstream location index, is:

$$\hat{\theta} = (\hat{\theta_1}, \hat{\theta_2}, \hat{\theta_3}, \hat{\theta_4}) = (30.5, 1.4, 5.4, 72.7)$$

From an hydrological point of view, the results confirm:

- a logical but quite weak increase in the upstream-downstream drift;
- a change, perhaps corresponding to a change in ditches function: from subsurface drainage with low flows on upstream areas to transport ditches with high flows.

We then calculated residuals $\hat{y}(s_\alpha) = z(s_\alpha) - m(s_\alpha)$.

Residuals Variogram Modelling and Estimation

We estimated $\gamma(h)$ on $y(\hat{s}_\alpha)$

Variogram Estimation

The variogram was estimated using the classical variogram estimator given in equation (7). The analysis was limited to 250 meters pairs distance on G_M. Figure 9 with points corresponding to experimental variogram computed on 10-m lags [5] shows:

- an order 2 stationarity of centered residuals $Y(s_\alpha)$ (**STA2**) with a bounded variogram;
- local dependencies, translated by an increasing variogram with small nugget effect which could be related to measurement error.

Variogram Fitting

An exponential variogram was fitted to the experimental variogram using a weighted least-squares method highly minimizing least squares at the shortest distances (Cressie, 1993, p. 95):

$$\gamma(h) = a\left(1 - \exp\left(-\frac{h}{b}\right)\right) + c$$

with $\theta = (a, b, c)$, corresponding to sill, range and nugget, respectively.
The fitted parameters are $\widehat{\theta} = (87.9, 705.8, 81.9)$.

Simulation of $Z(s) = m(s) + Y(s)$

We simulated fields for $Z(s)$ as is explained in appendix C. Simulated fields of $Y(s)$ were added to the drift values coming from the sub-model $m(s_\alpha)$ fitted above.

---

[5] numbers along the x abscissa correspond to the number of pairs for each lag.

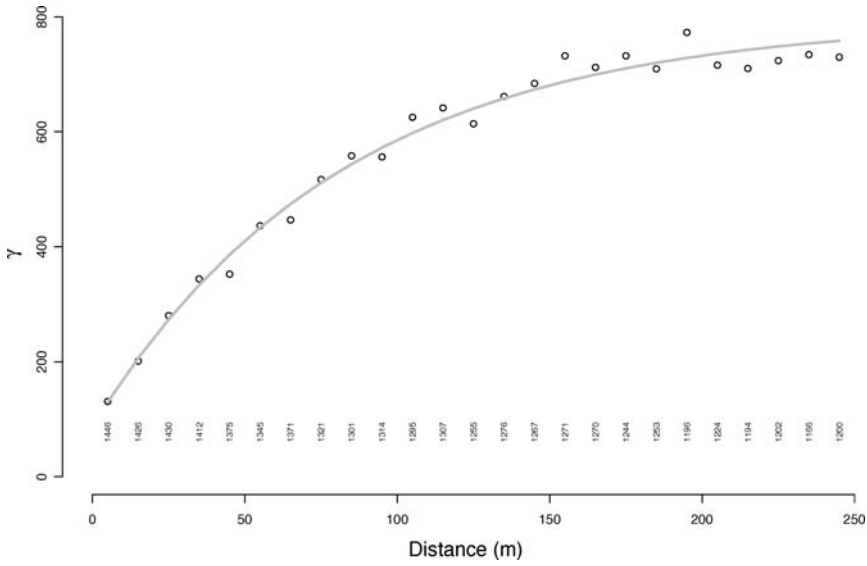

Figure 9. Experimental (points) and fitted (line) variograms on drift residuals.

Figure 10 (bottom) shows examples of simulated width values at the sampled locations. These simulations can be compared to the actual width distribution on the Roujan artificial drainage network shown the top of the Fig. 10.

These examples show realistic simulations with higher values located mainly in the center of the catchment and local variability along the ditches.

Computational Aspects

Drainage network geometry and topology were digitized in an arc topology GIS. Directed tree topological information and directed curvi-linear distances were computed within the GIS. We obtained the list of upstream-downstream curvilinear distances for all directed paths and the list of node degrees.

We developed routines used for data exploration, drift modelling and variogram estimation with the specific distances and metrics that we had defined, using the statistical free-ware R. Likewise the simulation procedure was also developed in the R framework, using the Cholesky decomposition in its Matrix package. Running the 500 simulations of 1500 values on a 62-river system (the Roujan artificial drainage network) took about 15 minutes with a 2 GHz processor and 1 GO RAM PC on windows system.

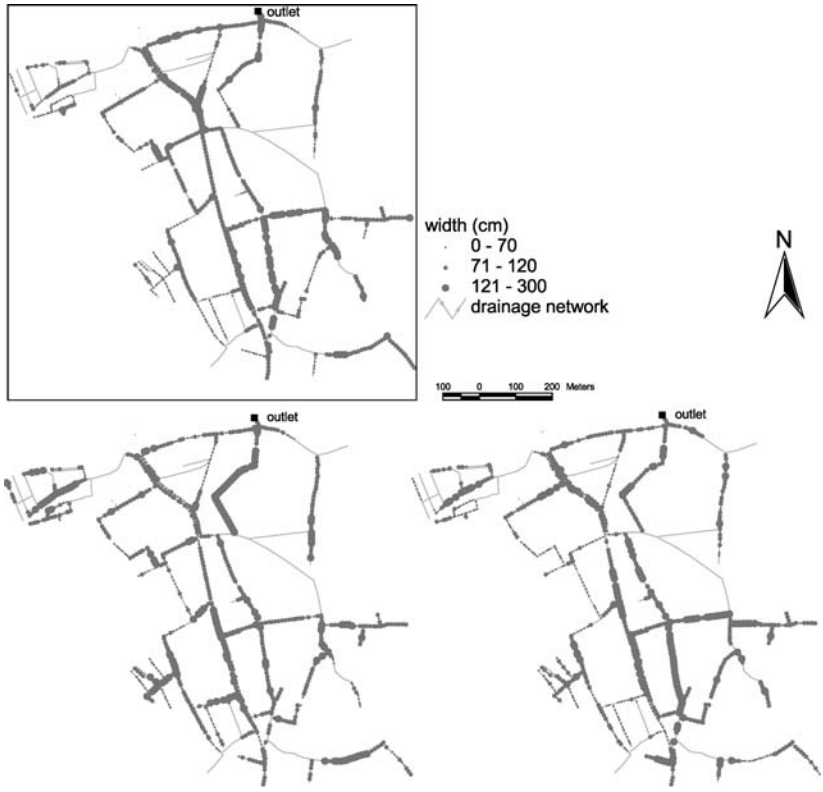

Figure 10. Width data on top-left depicted by dark grey points with legend on top-right. The 2 sub-figures down are examples of simulated values depicted with the same legend as for the actual ones.

DISCUSSION AND CONCLUSIONS

We have proposed a geostatistical model based on a random function defined at any point of the network, which can be used to simulate the spatial variability of local characteristics along a drainage network. We used the concept of a directed tree to model the specific features of drainage networks. This required reworking the classical geostatistical theory to suit the new framework.

Firstly we introduced the hypothesis of conditional independence between parts of the directed tree in order to define a new type of upstream-downstream stationarity and new variogram and covariance estimators. In this framework we proposed a procedure for simulating multi Gaussian fields on the network, that is suitable for water-flow related variables. This methodology was applied to study

width data of the Roujan artificial drainage network in Southern France. This case study highlights the advantages of using an approach based on fitted laws. Results exhibit a weak upstream-to-downstream drift but large spatial dependencies. The weak drift means that downstream and upstream sections are quite surprisingly similar. This indicates that the sections are not following the classical hydraulic model defined in Leopold and Maddock (1953). Strong spatial dependencies are not surprising and their range may result from the impact of human activities on the network.

Suggestions for Future Work

In this paper we have focused only on water-flow related variables that are continuous and normally distributed. Additional work would be required to simulate variables such as sediment cover that are not normally distributed. Furthermore, the local characteristics of recently created or well maintained artificial drainage networks could well be spatially distributed in homogeneous sections such as ditches. In that case, to reproduce reality more closely, the simulations would have to be able to generate homogeneous sections.

It might be difficult to accept the hypothesis of conditional independence between parallel parts of the network especially for variables governed by the adjacent land cover on cultivated catchments. Major work would be required to overcome this shortcoming. It would probably require a similar approach to the one proposed in this paper, starting out from the geostatistical basics under these new specific conditions.

Finally in this work we considered the geometry and topology of the network as being fixed. We would also have to work out to model and simulate these features in order to simulate the drainage networks for integration in distributed modelling.

ACKNOWLEDGMENTS

Authors are very grateful to Dr. Margaret Armstrong from the Ecole des Mines of Paris for help in careful editing of the paper.

REFERENCES

Audergon, J.-M., Monestiez, P., and Habib, R., 1993, Sampling dependences and sampling in fruit tree: a new concept for spatial prediction in fruit studies: J. Horticultural Sci., v. 68, no. 1, p. 99–112.

Beven, K. J., 1996, Response to comment on 'a discussion of distributed hydrological modelling by J.C. Refsgaard *et al.*', *in* Abbott, M. B., and Refsgaard, J. C. eds, Distributed Hydrological Modelling: Kluwer Academic Publishers, Dordrecht, p. 289–295.

Beven, K. J., 2001, How far can we go in distributed hydrological modelling: Hydrology and Earth System Sciences, v. 5, no. 1, p. 1–12.

Beven, K. J., and Feyen, J., 2002, The future of distributed modelling: Hydrological Processes, v. 16, p. 169–172.

Bouldin, J. L., Farris, J. L., Moore, M. T., and Cooper, C. M., 2004, Vegetative and structural characteristics of agricultural drainages in the Mississippi delta landscapes: Environmental Pollution, v. 132, no. 3, p. 403–411.

Carluer, N., and de Marsily, G., 2005, Assessment and modelling of the influence of man-made networks on the hydrology of a small watershed: implications for fast flow components, water quality and landscape management: J. Hydrology, v. 285, no. 1, p. 76–95.

Chilès, J. P., and Delfiner, P., 1999, Geostatistics : modeling spatial uncertainty: John Wiley and Sons, New York, 695 p.

Cressie, N., 1993, Statistics for spatial data: John Wiley and Sons, New York, 900 p.

Estupina Borrell, V., Llovel, C., and Dartus, D., 2002, Flash flood forecast hydrological model, in: Ninth International Conference on The Modelling, Monitoring and Management of Environmental Problems. Ed. Wessex Institute of Technology U., Bergen, Norway, p. 449–548.

Grayson, R., Blöschl, G., Western, A., and McMahon, T., 2002, Advances in the use of observed spatial patterns of catchment hydrological response: Advances in Water Resources, v. 25, no. 8–12, p. 1313–1314.

Guedon, Y., 2001, Pattern analysis in branching and axillary flowering sequences: J. Theoretical Biology, v. 212, p. 481–520.

Lagacherie, P., Diot, O., Domange, N., Gouy, V., Floure, C., Kao, C., Moussa, R., Robbez-Masson, J., and Szleper, V., 2005, An indicator approach for describing the spatial variability of human-made stream network in regard with herbicide pollution in cultivated watershed, *Ecological indicators*, v. 6, p. 265–279.

Lantuéjoul, C., 2002, Geostatistical Simulation, Models and Algorithms: Springer Verlag, Berlin, 256 p.

Legendre, P., and Legendre, L., 1998, Numerical ecology: 2nd English edition, Elsevier Science BV, Amsterdam, 853 p.

Leopold, L. B., and Maddock, T., 1953, The hydraulic geometry of stream channels and some physiographic implications, U.S. Geol. Surv. Prof. Pap. 252, p. 9–16.

Louchart, X., Voltz, M., Andrieux, P., and Moussa, R., 2001, Herbicides runoff at field and watershed scales in a Mediterranean vineyard area, Journal of Environmental Quality, v. 3, n. 30, p. 982–991.

Manly, B. J. F., 1997, Randomization, bootstrap and Monte-Carlo methods in biology, 2nd edition: Chapman and Hall, London, 399 p.

Moussa, R., Voltz, M., and Andrieux, P., 2002, Effects of the spatial organization of agricultural management on the hydrological behaviour of a farmed catchment during flood events: Hydrological Processes, v. 16, no. 2, p. 393–412.

Monestiez, P., Habib, R., and Audergon, J., 1989, Estimation de la covariance et du variogramme pour une fonction aléatoire à support arborescent: application à l'étude des arbres fruitiers, *in* Amstrong M. ed., Geostatistics, Kluwer Academic Publishers, Dordrecht, p. 39–56.

Raff, D. A., Smith, J. L., and Trlica, M. J., 2002, Statistical descriptions of channel networks and their shapes on vegetated hillslopes in Kemmerer: Wyoming, Hydrological Processes, v. 17, p. 1887–1897.

Ripley, B. D., 1987, Stochastic Simulation: John Wiley and Sons, New York, 237 p.

Varado, N., 2004, Contribution au développement d'une modélisation hydrologique distribuée. Application au bassin versant de la Donga au Bénin, 2004, Thesis, Grenoble University, 278 p.

Western, A., Blöschl, G., and Grayson, R., 2001, Towards capturing hydrologically significant connectivity in spatial patterns: Water Resources, v. 37, no. 1, p. 83–97.

APPENDIX A: DEFINITIONS AND NOTATION ON A DIRECTED TREE

A directed path L of size q is a sequence of q edges, $L = \{u_1, u_2, u_3, \ldots, u_q\}$ where each edge u_k of the sequence, for $2 \leq k \leq q-1$, has its initial extremity equal to the u_{k-1} final extremity and its final extremity equal to the u_{k+1} initial extremity.

The initial extremity i of u_1 and the final extremity j of u_q become respectively the initial and final extremities of path L.

The whole drainage network can be decomposed in a unique set of river $A_o = \{a_1, a_2, \ldots, a_{N_S}\}$, corresponding to the usual geographical denomination practice. The cardinal of A_o is the number of sources N_S.

An order on the tree vertices X is defined by sorting the path lengths between all vertices and the root. This order is then used to get an arbitrary decomposition of the whole tree in an ordered river subset of A. Let us define a list of N_S vertices, including r_G and all the junctions b_i, each one repeated $(d_G(b_i) - 2)$ times. The items are then sorted in a sequence following an ascending order for $l_{b_i r_G}$.

$$B_o = \left\{b_{(1)}, b_{(2)}, \ldots, b_{(N_S)}\right\} \text{ with } b_{(1)} = r_G \text{ and}$$
$$\forall (k, m) \in \{1, \ldots, N_S\}^2, k < m \Longrightarrow l_{b_{(k)} r_G} \leq l_{b_{(m)} r_G}$$

Let $A_o = a_{(1)}, a_{(2)}, \ldots, a_{(N_S)}$ be an ordered sequence of N_S rivers (Fig. 2B) which is sequentially defined by

$$a_{(1)} = \left(c_1, b_{(1)}\right) = (c_1, r_G) \text{ with } s_1 = \arg\max_{\{c|c\in S\}} \left(l_{cb_{(1)}}\right)$$
$$a_{(2)} = \left(c_2, b_{(2)}\right) \text{ with } c_2 = \arg \max_{\left\{c \left| \{c\in S-\{c_1\}\}\cap\{\Gamma_c \cap \Gamma_{c_1} \cap \Gamma^{-1}_{b_{(2)}} = b_{(2)}\}\right.\right\}} \left(l_{cb_{(2)}}\right)$$
$$\ldots$$
$$a_{(k)} = \left(c_k, b_{(k)}\right) \text{ with } c_k = \arg \max_{\left\{c|\{c\in S-\bigcup_{\alpha<k} c_\alpha\}\cap\{\forall \alpha<k, \Gamma_c\cap\Gamma_{c_\alpha}\cap\Gamma^{-1}_{b_{(k)}} = b_{(k)}\}\right\}} \left(l_{cb_{(k)}}\right)$$

A_o is a unique and arbitrary decomposition of the tree in rivers, each one associated with a source and a junction. The cardinal of A_o is the number of sources N_S. The order on junctions induces the order on the rivers.

APPENDIX B: CONDITIONAL INDEPENDENCE FOR GEOSTATISTICAL MODELS ON A DIRECTED TREE

Conditional independence between parallel rivers is defined by the equality of two conditional distributions:
$\forall n \in \mathbb{N}, \quad \forall s_1, \ldots, s_n \in G_M, \quad \forall s' \in G_M,$ the set of s_i's is arranged in three subsets reordering the indices so that $s_1, \ldots, s_\alpha$ belong to $\Gamma^{-1}_{s'}$, $s_{\alpha+1}, \ldots, s_\beta$ belong

to $\Gamma_{s'}$ and $s_{\beta+1}, \ldots, s_n$ the others s_i, are located on parallel rivers to the one s' belongs to. Let $Z(\Gamma_{s'})$ denote the random function Z for all points that belong to $\Gamma_{s'}$ including $s_{\alpha+1}$ to s_β. Then:

$$\mathcal{L}(Z(s_1), \ldots, Z(s_\alpha)|Z(s_{\beta+1}), \ldots, Z(s_n), Z(\Gamma_{s'})) = \mathcal{L}(Z(s_1), \ldots, Z(s_\alpha)|Z(\Gamma_{s'})) \quad (9)$$

If the realization of the random function downstream s' is known, any other information on what happens on parallel rivers is useless to characterize what happens on locations that are upstream s'.

The stationarity is then simply defined for any path L by:

$\forall n \in \mathbb{N}, \quad \forall z_1, \ldots, z_n \in \mathbb{R}^n,$

$\forall L, \quad \forall s_1, \ldots, s_n \in L$ and $\forall s'_1, \ldots, s'_n \in L$ such as $s'_1 \in \Gamma_{s_1}, \ldots, s'_n \in \Gamma_{s_n}$

and such as $d(s_1, s'_1) = h, \ldots, d(s_n, s'_n) = h$

$$P(Z(s_1) < z_1, \ldots, Z(s_n) < z_n) = P(Z(s'_1) < z_1, \ldots, Z(s'_n) < z_n) \quad (10)$$

that corresponds to an invariance by upstream-downstream translation in a geometric space of dimension one.

In consequence, the finite dimensional distribution on $s_1, \ldots, s_n$ can be always expressed as products of distributions on subsets of s_i's belonging to edges from a same path L.

APPENDIX C: SIMULATION OF RANDOM FUNCTION ON A DIRECTED TREE

Let $\mathcal{D} = \{s_1, \ldots, s_\alpha, \ldots, s_N\}$ be the finite set of points of G_M where a random simulation of $Y(s_\alpha)$ is needed.

We first decompose $\mathcal{D}$ in subsets of points such that each site is associated with the river from A_o to which it belongs.

After ordering the sites according to the river sequence $a_{(1)}, a_{(2)}, \ldots, a_{(N_S)}$ of A_o, we denote $\mathcal{D}$ using new subscripts

$$\mathcal{D} = \left\{s^1_{(1)}, \ldots, s^1_{(n_1)}, s^2_{(1)}, \ldots, s^2_{(n_2)}, \ldots, s^{N_S}_{(1)}, .., s^{N_S}_{(n_{N_S})}\right\}$$

where the subset $\{s^k_{(1)}, \ldots, s^k_{(n_k)}\}$ is the subset of all sites that belong to the k^{th} river $a_{(k)}$ of A_o.

Step 1: Main River $a_{(1)}$ Simulation

The simulation starts on $a_{(1)}$ river, with the subset of points $\{s^1_{(1)}, \ldots, s^1_{(n_1)}\}$ belonging to $a_{(1)}$.

Let $Y_{a_{(1)}}$ denote a random vector $Ys^1_{(1)}, \ldots, Ys^1_{(n_1)}$. Its distribution is multi-Gaussian with mean 0 and covariance $\Sigma_{(0)}$. $\Sigma_{(0)}$ is calculated from the covariance function $C(d(s_\alpha, s_{\alpha'}))$ where s_α and $s_{\alpha'} \in \{s^1_{(1)}, \ldots, s^1_{(n_1)}\}$.

A way to simulate $Y_{a_{(1)}}$ is:

$$Y_{a_{(1)}} = L_{(1)}\varepsilon_{(1)} \tag{11}$$

where $\varepsilon_{(1)}$ is a random vector of length n_1 made of i.i.d. Gaussian random variables $\mathcal{N}(0, 1)$ and $L_{(0)}$ results from the Cholesky decomposition of $\Sigma_{(0)} = L_{(0)}.{}^tL_{(0)}$.

Any other procedure to simulate a multi-Gaussian distribution for $Y_{a_{(1)}}$ can be used, in particular if its size n_1 becomes very large and Cholesky decomposition problematic.

Step 2: Iterative River Simulations on the Ordered Series *Ao*

The problem is now to simulate the $Y(s_\alpha)$ values for points belonging to any $a_{(k)}$ conditionally to already simulated values on $\{a_{(1)}, \ldots, a_{(k-1)}\}$.

For a given k, let us define three subsets of $\mathcal{D}$.

The first one D_1 contains the points to simulate $D_1 = \{s^k_{(1)}, \ldots, s^k_{(n_k)}\}$, the second D_2 the points already simulated $D_2 = \{s^1_{(1)}, \ldots, s^1_{(n_1)}, \ldots, s^{k-1}_{(1)}, \ldots, s^{k-1}_{(n_{k-1})}\}$ and the third $D_3 = \bigcap_{s\in D_1} \Gamma_s^{-1} \cap D_2$ that is the intersection between the set D_2 of already simulated points and the set of downstream points of the ones to be simulated.

From Equation (9), it follows for the following distributions:

$$(Y(s_\alpha), s_\alpha \in D_1 | \, Y(s_\nu), s_\nu \in D_2) \sim (Y(s_\alpha), s_\alpha \in D_1 | \, Y(s_\beta), s_\beta \in D_3)$$

and only the joint distribution on D_1 and D_3 is needed.

Let $\Sigma_{(k)}$ be the covariance matrix of points $s_\alpha \in D_1 \cup D_3$ defined by $C(d(s_\alpha, s'_\alpha))$.

Then, in a manner similar to the conditional simulation method proposed by Journel and Huijbregts in Cressie (1993), p. 207., we define

$$\forall s_\alpha \in a_{(k)}, \quad Y(s_\alpha) = Y_{nc}(s_\alpha) + \left(Y^*(s_\alpha) - Y^*_{nc}(s_\alpha)\right) \tag{12}$$

where

- $Y^*(s_\alpha)$ is the predicted value of $Y(s_\alpha)$ by simple kriging, using as data set the values of $Y(s_\beta)$ for $s_\beta \in D_3$ and as covariance matrix $\Sigma_{(k)}$.
- $Y_{nc}(s_\alpha)$ is obtained by an independent simulation of all points in $D_1 \cup D_3$ using the same simulation algorithm as in step 1 with $\Sigma_{(k)}$ as covariance matrix.
- $Y^*_{nc}(s_\alpha)$ is the predicted value of $Y(s_\alpha)$ by simple kriging, using as data set the simulated values $Y_{nc}(s_\beta)$ for $s_\beta \in D_3$.

It is straightforward to check that the vector $\big(Y(s_1), \ldots, Y(s_{n_k})\big)$ on $a_{(k)}$ follows correct multivariate distribution and that the joint distribution with the $Y(s_\beta)$ on already simulated downstream points honors the $C(h)$ spatial covariance structure. Step 2 is successively applied until the last river $a_{(N_S)}$ of A_o. The simulation of all points of $\mathcal{D}$ is then complete.