A spatial stochastic algorithm to reconstruct artificial drainage networks from incomplete network delineations

J.S. Bailly, F. Levavasseur, P. Lagacherie

A spatial stochastic algorithm to reconstruct artificial drainage networks from incomplete network delineations

J.S. Bailly[a,c,*], F. Levavasseur[b], P. Lagacherie[b]

[a] *AgroParisTech, UMR LISAH, F-34060 Montpellier, France*
[b] *INRA, UMR LISAH, F-34060 Montpellier, France*
[c] *AgroParisTech, UMR TETIS, F-34093 Montpellier, France*

Abstract

A spatial stochastic algorithm that aims to reconstruct an entire artificial drainage network of a cultivated landscape from disconnected reaches of the network is proposed here. This algorithm uses random network initialisation and a simulated annealing algorithm, both of which are based on random pruning or branching processes, to converge the multi-objective properties of the networks; the reconstructed networks are directed tree graphs, conform to a given cumulative length and maximise the proportion of reconnected reaches. This algorithm runs within a directed plot boundaries lattice, with the direction governed by elevation. The proposed algorithm was applied to a 2.6-km^2 catchment of a Languedocian vineyard in the south of France. The 24-km-long reconstructed networks maximised the reconnection of the reaches obtained either from a hydrographic database or remote sensing data processing. The distribution of the reconstructed networks compared to the actual networks was determined using specific topographical and topological metrics on the networks. The results show that adding data on disconnected

---

[*]Corresponding author
  Email addresses: `bailly@teledetection.fr` (J.S. Bailly),
`levavasseur@supagro.inra.fr` (F. Levavasseur), `lagacherie@supagro.inra.fr`
(P. Lagacherie)

reaches to constrain reconstruction, while increasing the accuracy of the re-constructed network topology, also adds biases to the geometry and topography of the reconstructed network. This network reconstruction method allows the mapping of uncertainties in the representation while integrating most of the available knowledge about the networks, including local data and global characteristics. It also permits the assessment of the benefits of the remote sensing partial detection process in drainage network mapping.

Keywords: channels, ditches, mapping, graphs, simulated annealing, simulation, uncertainties, remote sensing, random walks

---

1. Introduction

Artificial drainage networks of agrarian landscapes include connected linear features, such as channels, culverts, roads and reshaped gullies, that result from the constant efforts of farmers to adapt landscapes to the constraints of agricultural production. They constrain water flow paths in landscapes, which greatly impacts the environmental functioning of agrarian areas: drainage networks can be ecological hot-spots and corridors (Carpentier et al., 2003; Croxton et al., 2005; Pita et al., 2006) and control many fluxes, especially of water and pollutants (Bouldin et al., 2004; Dages et al., 2008).

When scaling up from an agricultural plot scale, the structure of drainage networks appears to be highly variable in space (Lagacherie et al., 2006). Therefore, most environmental diagnoses and impact assessments at the catchment scale in agrarian areas now take these networks and their spatial variabilities into account through physical modelling (Lagacherie et al., 2010; Moussa et al., 2002; Varado, 2004). However, to date, these approaches have only been used for small areas where major investments were made in terrain surveys to map drainage networks (Lagacherie et al., 2006), which limits the

2

generalisation of these studies. Consequently, the extension of environmental agrarian landscape diagnoses to wider areas with a reasonable cost is largely dependent on our ability to map such linear elements.

Similar to most surface linear mapping methods (Amini et al., 2002), remote sensing represents a candidate method for artificial drainage network mapping (Vassilopoulos et al., 2008). Methods allowing the delineation of drainage networks in natural areas from regular grid digital terrain models (DTMs) have existed since 1984 (O'Callaghan and Mark, 1984); these methods use drainage algorithms to search for the flow direction in each pixel (Wilson et al., 2007) and build aggregated drainage areas accordingly (e.g., vectorised thresholded drainage areas usually form the final drainage networks) (Soille and Gratin, 1994).

For cultivated landscapes, drainage algorithms used on DTMs are unable to represent anthropogenically modified overland flow paths (Duke et al., 2006; Garcia and Camarasa, 1999), even when using high resolution airborne scanner laser DTMs (Comoretto, 2003).

Instead, methods delineating network reaches from images extracted from whole networks have previously been assessed (Bailly et al., 2008; Vassilopoulos et al., 2008). However, as for most raw image processing results, confusion in remote sensing detection provides an incomplete view of a network by producing sets of unconnected and uncertain reaches.

Consequently, methods that aim to reconstruct an entire likely connected drainage network from an incomplete delineated network must be proposed. These methods could be used to reconnect reaches detected by remote sensing, as well as to extend known downstream regions of networks existing in geographical databases. In the case of mapping explicit uncertainties, these methods might allow assessment of the ultimate utility of specific remote

sensing processes for network mapping or the required amount of known downstream reaches of networks in geographical databases. Furthermore, with regard to the use of drainage network maps for modelling the geometry of fluxes (Beven and Binley, 1992; Beven and Freer, 2001), methods are also needed that represent uncertainties in network reconstruction, i.e., uncertainties in network mapping.

Methods that aim to reconstruct line networks (e.g., roads and hydrographic networks) from incomplete network delineations from remote sensing (Stoica et al., 2004; Lacoste et al., 2005) or vectorised cartographic elements (Baltsavias and Zhang, 2005; Mariani et al., 1995) have previously been proposed. The former method uses a random polyline stochastic simulation process derived from the Candy model (Descombes et al., 2001), and the latter uses a deterministic minimisation process (Kruskal, 1956; Prim, 1957). Both reconnect networks with geometric rules, favouring long, slighty curved polylines without overlapping or aligned segments. However, at the end of the process, both methods provide a single estimated network, without computation of uncertainties.

Considering this last point, we chose first to develop a reconstruction algorithm to: 1) randomly reconstruct various possible artificial drainage networks from an incomplete delineation of the network and 2) reconstruct networks taking into account a number of global network geometrical and topological properties. The algorithm employed here is a spatially simulated annealing algorithm that was chosen for its ability both to better explore all network possibilities and to escape from local minima during convergence (VanGroenigen and Stein, 1998). This algorithm generates networks corresponding to directed trees (Fairfield and Leymarie, 1991) and reconnecting reaches of the incomplete delineated network. The reconstructed networks

4

correspond to connected sub-graphs of the directed lattices of agricultural plot boundaries. The proposed algorithm falls within the scope of spatial conditional simulation methods (Lantuejoul, 2002). The sampled reaches of the network conditioning the simulation can come from ground observations, remote sensing detection or existing geographical databases. The direction of the graph comes from elevation gradients. Uncertainties in mapping networks are simulated through a set of equi-probable generated networks.

To stringently evaluate how these uncertainties impact network morphology in a catchment space, methods measuring how similar the reconstructed networks are to the actual networks need to be employed. In network mapping, this is usually carried out using only geographical overlap criteria (Molly and Stepinski, 2007; Heipke et al., 1997). However, this method does not permit the assessment of how realistic the morphology of the reconstructed network is. Thus, specific complementary similarity metrics of networks were developed for this purpose.

This paper is organised into four sections: 1) a description of the developed methods, i.e., the reconstruction algorithm (its assumptions, principles and general parameters) and the network topological and and topographical dissimilarity metrics are first exposed; 2) a case study on the drainage ditch network of a $2.6 - km^2$ vineyard catchment located in the south of France, with different sets of data conditioning the network reconstruction and two simulated annealing convergence criteria being used for network reconstruction is secondly presented; 3) comparisons of topological and and topographical dissimilarities between the actual network and some reconstructed networks are performed; and 4) we conclude and expand on the limits of the proposed method to properly assess the implications of the findings produced from an incomplete delineation network from remote sensing.

5

2. Materials and methods

2.1. The reconstruction algorithm

An artificial drainage network (ADN) is considered to be a physical network corresponding to a directed and valued graph d with an edge value equal to the reach length of an according physical network and a vertex value equal to the elevation. Using the usual graph theory notations, $d = (x_d, u_d)$ where x_d denotes the set of d vertices and u_d the set of d edges. An edge corresponds to a directed or ordered pair of vertices (n_i^+, n_j^-), where n_i^+ denotes n_i as the start-vertice and n_i^- denotes n_i as the end-vertice.

2.1.1. Assumptions

We first considered that the graph d that we want to simulate is a connected sub-graph of the valued, connected, planar lattice representing the cultivated plot boundaries within a cultivated landscape. ADN reaches are, in fact, almost always located on plot boundaries. This lattice is denoted similarly as $l = (x_l, u_l)$.

l is a directed graph (digraph), i.e., a directed lattice. The direction is given by the elevation gradient between edge vertices; it corresponds to the upstream-downstream direction. However, considering that elevation is uncertain due to the noise σ_z parameter, an ambiguous direction for a given plot boundary ending on the two vertices n_i, n_j can be obtained. In this case, we chose to retain the two edge directions, which meant that two distinct edges, v and v', were considered possible between vertices n_i, n_j:

$$\forall n_i, n_j \in x_l^2 /_{|z_{n_i} - z_{n_j}| < \sigma_z}, v = (n_i^+, n_j^-) \in u_l, v' = (n_j^+, n_i^-) \in u_l \qquad (1)$$

6

As has been done in most ADN hydrology studies (Lagacherie et al., 2010; Moussa et al., 2002; Varado, 2004), we considered that the d graph is also a directed tree (ditree) with a root r corresponding to the physical outlet.

To assimilate any type of existing data for ADN locations, we also consider that an unconnected sub-graph exists: $f = (x_f, u_f) \in d$. u_f denotes a set of *fixed* edges. This can be produced from an existing geographical database related to a hydrographic network for downstream reaches, from a remote sensing detection process, or from any field survey information. As an extremal case, f can only contain the r vertex, the fixed outlet. Additionally, the subsets of fixed edges u_{f_1} and u_{f_2}, such as $u_{f_1} + u_{f_2} = u_f$, are also distinguished. u_{f_2} denotes the set of fixed edges with a fixed direction (typically obtained from a geographical database), regardless of the slope gradient. u_{f_1} denotes the set of fixed edges with an uncertain direction (typically obtained from a remote sensing detection process).

2.1.2. Spatial simulated annealing algorithm general scheme

Simulated annealing is carried out for the optimisation of stochastic algorithms based on the Metropolis-Hastings (MH) heuristic. Basically, it allows the optimisation of paths through costly points when the temperature is still high in the first iterations of the algorithm. It is an heuristic that is usually employed for spatial problems, and it permits a better exploration of space from an initial solution (VanGroenigen and Stein, 1998).

The proposed formulation of the algorithm can be written in pseudo-code as follows:

In this algorithm, the parameters are the generally employed simulated annealing parameters: k_{max} is the maximum number of iterations; e_{max} is

---

Algorithm 1 Reconstruction digraph algorithm

1: **procedure** SIMULATED.ANNEALING($d_0, e_{max}, k_{max}, T_{ini}, g$)

2:     $k \leftarrow 0$

3:     $d \leftarrow d_0$ ▷ initial state

4:     $e \leftarrow energy(d_0)$ ▷ initial energy

5:     $T \leftarrow T_{ini}$ ▷ initial temperature

6:     **while** $k < kmax$ & $e > e_{max}$ **do**

7:         $d_{new} \leftarrow neighb(d)$ ▷ random neighbour state

8:         $e_{new} \leftarrow energy(d_{new})$ ▷ energy function

9:         **if** $e_{new} < e$ **then**

10:             $d \leftarrow d_{new}; e \leftarrow e_{new}$

11:         **else if** $\exp\left(\frac{e-e_{new}}{T}\right) > random(U[0,1])$ **then**   ▷ MH heuristic

12:             $d \leftarrow d_{new}; e \leftarrow e_{new}$

13:         **end if**

14:         $k \leftarrow k+1$

15:         $T \leftarrow g(T,k)$ ▷ Temperature decreasing law

16:     **end while**

17:     **return** d ▷ Return reconstructed digraph

18: **end procedure**

---

the energy objective criteria to reach; and T_{ini} is the initial temperature. $g(T,k)$ denotes a given decreasing law of temperature. d_0 corresponds to the initial state for digraph reconstruction. The $random(U[0,1])$ function gives a sample from the uniform distribution.

At the end of the algorithm 1, a digraph d connected to root r is obtained satisfying the e_{max} criterion. Due to its construction, few cycles can occur in this digraph. To ensure a final ditree structure for d, the cycles are broken

in a final deterministic uncycling process that disconnects the longest paths from the root (in edges number) and guarantees continuity in successive u_{f_2} edges.

By replicating this stochastic reconstruction process, a forest of n reconstructed ADN, denoted $(d_1, \ldots, d_n)$, can be simulated. This forest represents uncertainties in ADN mapping.

In the following, the initialisation d_0, the production of a random neighbour state $neighb$, the decreasing temperature law and the definition of the $energy$ function are successively detailed. Prior to this, the random walk algorithm, on which the two former algorithms are based, is briefly presented.

2.1.3. Random walk

Here, a random walk process consists of random sequential and independent paths in a directed lattice starting from an upstream vertex and moving down to a set of downstream end vertices (Spitzer, 1970).

2.1.4. Random branching initialisation

A random initial state for a digraph d_0 is obtained using a sequential branching process based on random walks within lattice l. The process begins from the single graph r. At each iteration based on ordered x_f vertices related to ascending elevation, the process connects the x_f considered vertex to the digraph obtained in the previous iteration by a random walk. The digraph grows to obtain a given ratio of u_f edges connected to the root r.

2.1.5. Neighbour state process: random pruning and branching

During simulated annealing, neighbouring random states for digraph d are generated (see line 7 in algorithm 1). A neighbour state is based on

random pruning (Fig. 1-b) and branching (Fig. 1-a) of the d graph as depicted on figure 1 on a virtual example.
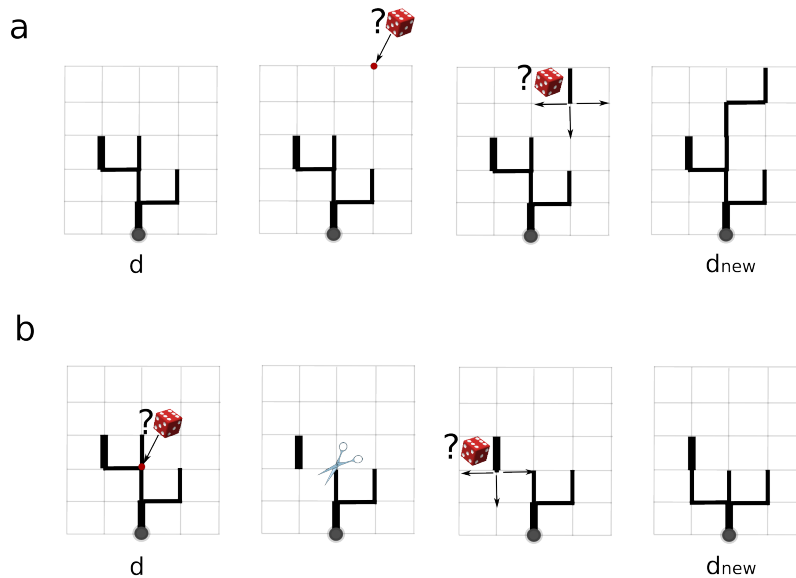

Figure 1: A neighbour state of tree d generated through random branching (a) or random pruning (b) processes. The two wide black segments are fixed parts of the network and the wide grey point corresponds to the tree root (outlet).

In this process, an edge belonging to the $(u_l - u_{f_2})$ set is first randomly selected (selection among the uncertain reaches), and two cases are therefore distinguished:

- If this selected edge already belongs to the u_d digraph set of edges, then the directed sub-graph (branch) upstream of the ending vertex of this edge is first pruned from digraph d to form d_{new}. Second, a re-branching process is selected through a binary process with a probability of 0.5. If selected, this re-branching process connects each of

the u_f edges belonging to the upstream sub-graph with same sequential branching process as was used for the initialisation of d_0, and this new branch is added to d_new.

- If the selected edge does not belong to (u_d), a random walk starting from its end node, reaching one of the d vertices, is performed. Consequently, a new branch is added to form d_{new}.

Due to its construction, this neighbour process allows the exploration of the entire l lattice for the reconstruction of d.

2.1.6. Energy functions

The choice of the energy function in the algorithm1 governs the objective properties for the reconstructed networks. Here, we chose a bi-objective *energy* function. Our aim was that a reconstructed network should:

- converge to a target overall cumulative network length related to the extent of the catchment: this defines the catchment objective *catch* ; by using this convergence criterion, it is assumed that in future cases, the drainage network density or the cumulative length of a catchment can be predicted from environmental or anthropogenic variables (e.g., soil, catchment morphology and cultivated plot areas. . .).

- reconnect a maximum number of fixed edges at the extent of the catchment: this defines the reconnection ratio objective *ratio*, equal to the fraction between the number of reconnected fixed edges and the number of fixed edges;

The multi-objective energy function is $energy = max(e_1, e_2)$, where e_1, e_2 denote component energy objective functions:

catchment energy: $e_1 = \frac{|L_d - catch|}{\sigma_1}$ with $\sigma_1 = catch * \gamma_1$

reconnection ratio energy: $e_2 = \frac{|L_{f \in d} - L_f|}{\sigma_2}$ with $\sigma_2 = L_f * \gamma_2$

In these formulations, $L_d = \sum_{v \in u_d} length(v)$, $L_f = \sum_{v \in u_f} length(v)$,

$L_{f \in d} \sum_{v \in (u_d \& u_f)} length(v)$ denotes the cumulative length of f, d and f,

respectively, within d edges. γ_1, γ_2 are convergence tuning parameters that

permit the differential adjustment of each of the objectives (e_1, e_2) relative

to the unique e_{max} parameter.

Due to the construction of these formulations, all of the energies have

normalised magnitude ranging from 0 to 1.

2.1.7. Simulated annealing decreasing temperature law

The simulated annealing initial decreasing temperature law follows the

commonly employed equation:

$$T_k = T_{ini} * (\alpha^k) \tag{2}$$

where:

- $\delta = max(1, \frac{(L_l - catch)}{catch} * 0.5)$

- $T_{ini} = \frac{\delta}{0.51}$

The initial temperature parameter δ was fixed to a 'great' cost, divided

by 2 and normalised by the target catchment cumulative length. The α

parameter was fixed such that $\exp\left(\frac{-0.2}{(\alpha^{(\frac{k_{max}}{5})}) * 5}\right)$ approximates 0.1. Both

0.51 and 0.1 values were fixed arbitrarily after trials and errors tests since

automated fitted parameters for simulated annealing is still an open question

(Wright, 2010). There were also fixed according to suggestions from Aarts

et al. (2003) for initial temperature related to the maximal cost between

12

two neighbouring solutions and from Zhang et al. (2010) in order to get a
rejection rate of about 5-10 %.

2.2. Evaluating the ADN simulations with drainage network similarity metrics

The previous subsection showed the way an ADN would be reconstructed
given the following information: r, the known outlet of the network; l, the
plot boundary lattice graph valued by elevation, and f, the set of fixed
network reaches used to condition the reconstruction.

To assess the accuracy of the proposed ADN mapping method with respect to the f data or algorithm parameters, we computed the ditree network
similarities based on scalar metrics. Theses metrics were used to compare
each of the simulated ADNs to the actual one and, consequently, the forest
dissimilarity distribution was computed.

As the objective function in algorithm 1 already uses global network
geometry criteria (the drainage density), we chose to develop two complementary network similarity metrics based on: 1) network topography and
2) network topology.

2.2.1. ADN topographical similarity metric

The topography of an ADN is fully contained in the slope of its network
reaches. To summarize these slope distributions, any single synthetic slope
statistic (e.g., mean and quantile) performs poorly. Therefore, we chose to
use a metric based on the cumulative distribution of edge slopes. For a given
network, we calculated the curve representing the cumulative distribution
of slopes as follows: 1- for a given slope value θ, the proportion of the
network length having slope smaller or equal than θ was computed; 2- this

was repeated for discretized and crescent values of the slope ranging from -90 to 90 degrees (see Figure 2).

For two networks, the topographical similarity corresponds to the area between the two curves. This metric h_t can be defined as:

$$h_t(d_1, d_{ref}) = |\int_{-90}^{90} F_i(\theta)d\theta - \int_{-90}^{90} F_{ref}(\theta)d\theta|, \tag{3}$$

where θ denotes the slope in degrees and $F_i(\theta)$ denotes the empirical cumulative distribution function of the network slopes in relative cumulative lengths.
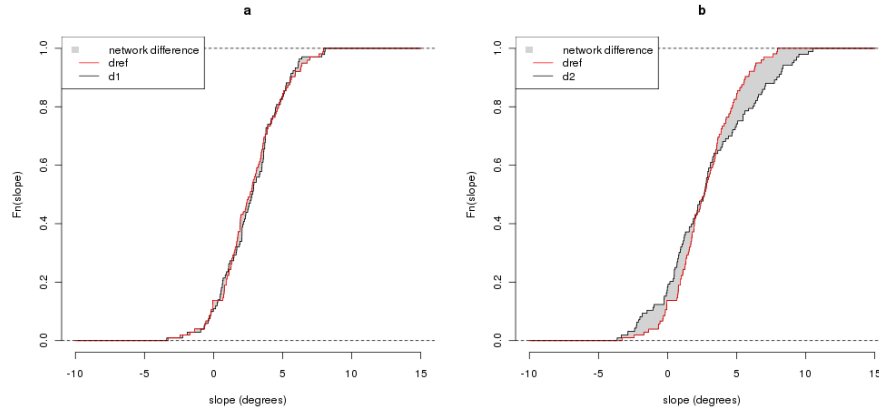

Figure 2: ADN topographical similarity computation through empirical cumulative edge slope distribution difference (grey areas): example of small (a) and great (b) dissimilarities for ditree d_1 and d_2 compared to d_{ref}

2.2.2. ADN topological similarity metric

Methods for the computation of the similarity between tree topologies have previously been developed in computational science to answer prob-

lems related to character chains, phylogenetic tree comparisons or the description of deciduous tree architecture (Ferraro and Godin, 2000). These methods are all based on the Robinson-Foulds metric (Day, 1985) or Levenshtein or Jaro-Winkler distances (Jaro, 1985). These metrics are dedicated to valued trees with qualitative node labels. However, this is clearly not the problem in our case. In hydrology, topological characterisation of drainage networks is usually based on the Horton-Strahler, Shreve or Tokunaga upstream-downstream edge or node taxonomies (Barker et al., 1973). However, for a given ditree, these taxonomies cannot be easily reduced to a single value: it is usually represented in vectors or matrices as a branching matrix based on bi-order distributions, i.e. distributions of the twostrahler orders for stream joining downstream (Janey, 1992, chapitre3).

To address this problem, we chose to develop the simple similarity metric h_c, as proposed initially by Moslonka-Lefebvre et al. (2010) for comparison of gully networks in badlands extracted from remote sensing data (Thommeret et al., 2010). This metric is the distance between the Strahler barycenters of networks, where the Strahler barycenter of a network is the weighted barycenter of the network edge upstream nodes, with weights equal to Strahler orders:

$$h_c(d_1, d_2) = distance(G_1, G_2) \tag{4}$$

- $x_{G_1} = \frac{1}{\sum_i w_{1_i}} \sum_i w_{1_i} x_{1_i}$; $y_{G_1} = \frac{1}{\sum_i w_{1_i}} \sum_i w_{1_i} y_{1_i}$ are the coordinates of the d_1 Strahler barycenter tree;

- $x_{G_2} = \frac{1}{\sum_j w_{2_j}} \sum_j w_{2_j} x_{2_j}$; $y_{G_2} = \frac{1}{\sum_j w_{2_j}} \sum_j w_{2_j} y_{2_j}$ are the coordinates of the d_2 Strahler barycenter tree;

15

- w_{1_i}, w_{2_j} denotes the Strahler order of a node equal to that of the edge for which the node is the upstream for trees d_1, d_2 respectively. x_{1_i}, x_{2_j} and y_{1_i}, y_{2_j} denote the coordinates of the edge upstream vertices of trees d_1, d_2 respectively.

This unique metric combines the topology, geometry and geography of the network. It is further simply considered as the topological similarity metric.

3. Case study

3.1. The Muscat catchment and drainage network

3.1.1. Study site

The study area used here is the 2.6-km^2 Muscat catchment located in the Languedoc region (south of France). This area was selected as being representative of the Languedocian vineyard landscape, where linear elements play a major role in hydrological transfers. The area is successively composed of plateaus, sloping areas with terraces, gentle foot slopes formed from alluvial and colluvial deposits and a flat depression from upstream to downstream. The study area includes the Roujan elementary catchment, which has been equipped for conducting measurements of pollution, erosion and flooding since 1992 (Voltz and Albergel, 2002). The drainage network drains cultivated plots and prevents erosion in the case of major runoff events. The positioning of the reaches that comprise most of this network within the landscape is highly variable and is highly dependent on reach functions. When the reach function is to capture surface waters, as in the case of high slopes, they are perpendicular to the terrain slope, especially

16

at the footstep of terrassettes. When the reach function is water transport, the reaches follow the slope direction. When reach function is to drain subsurface waters, there is no evident relationship with the local relief.

3.1.2. The actual Muscat drainage network

The actual drainage network of the studied catchment was exhaustively surveyed in 2004. It contains 239 edges and 240 nodes with 123 sources and 116 confluences (Fig. 3-a). Its cumulative length is 23,995 m that defines the *catch* parameter value.
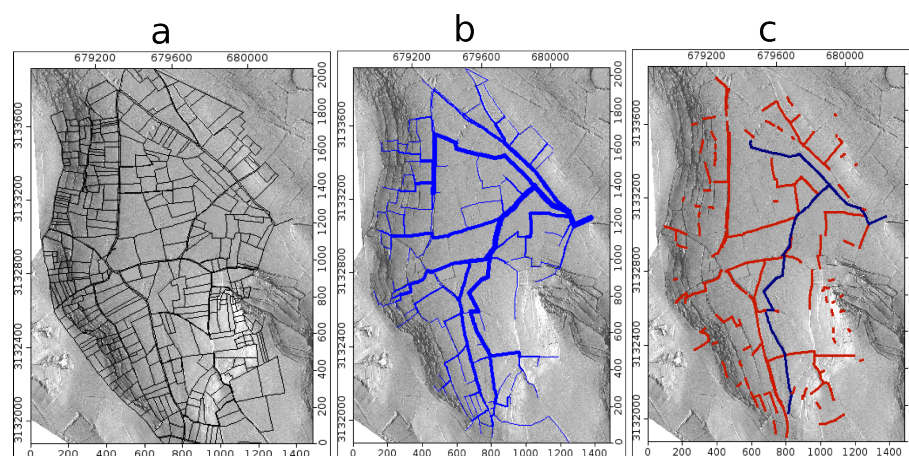

Figure 3: The Muscat catchment data: a) the cultural plot lattice depicted on the 5-m hillshaded DTM; b) the actual drainage network in blue, with the reach width related to Strahler order ranging from 1 to 5; c) the plot lattice with the set f of known edges (blue: © BD-Topo for f_2, red: remote sensing for f_1)

3.1.3. The plot lattice and DTM data

For this catchment, a cultivated connected plot lattice containing 2,645 edges and 1,079 nodes was used (Fig. 3-b). The elevation was assigned to

17

each vertex of this lattice based on a 5-m resolution DTM having a random error of $\sigma_z = 1m$ for the standard deviation. There is no internal sink within this lattice that can trap random directed walks.

3.1.4. The ADN conditioning data

We used the whole set of known edges, f, to condition the reconstruction, as depicted in Figure 3-c. This set was produced from:

- remote sensing detection based on LIDAR raw data (red lines). Remote sensing provides disconnected reaches, which may represent false detection, with a probability of 0.17 of false detection for upstream reaches that were not previously mapped in a geographical database. The cumulative length of the reaches is 15,400 m, i.e., 47% of the actual network cumulative length. This set of fixed reaches corresponds to the f_1 edge set because the direction of the reaches is unknown.

- the © BD-Topo-geographical database for hydrographic networks for the main downstream reaches (blue). Their cumulative length is 2,500 m, i.e., 7.5% of the actual network. This set of fixed reaches corresponds to the f_2 edge set, which has a known direction.

- r which is the known outlet of the network that we want to reconstruct. Although it is not appropriate (no edge), we further consider that r can be the extreme case for the f conditioning data: only the outlet is fixed.

4. Results: reconstructed ADN on the Muscat catchment

4.1. Convergence efficiency

Beginning from the cultivated plot lattice shapefiles, algorithms were developed in the R statistical computing program (Ihaka and Gentleman, 1996) using the *shapefiles* and *e1076* libraries (Team, 2009). It took 9, 7.2 and 48 minutes on average to reconstruct a network using the $(f = r)$, $(f = f_2)$ $(f = f_1 + f_2)$ sets for conditioning, respectively, on a dual core PC with 1 GB RAM. Additionally, r, f_2, $f_1 + f_2$ denote the three exclusive cases of f fixed edges used to condition the reconstruction.

The $e_{max}, \gamma_1, \gamma_2, k_{max}$ parameters were adapted depending on the f conditioning data:

- when using r or f_2 to condition the reconstruction, the convergence parameters were $\gamma_1 = 1$, $\gamma_2 = 1$, $e_{max} = 0.01$ and $k_{max} = 500$.

- when using $f_1 + f_2$ to condition the reconstruction, it took a long time to achieve convergence due to both considerable constraints related to lattice space exploration and a multi-objective (e_1, e_2) dilemma (Kung et al., 1975). A Pareto optimum providing a compromise between the duration of and the objective of accuracy for the simulation was found with the following convergence parameters: $\gamma_1 = 1$, $\gamma_2 = 4$ for $e_{max} = 0.05$ and $k_{max} = 2,000$.

In Figure 4, the evolution of energy within the simulated annealing process using $f = (f_1 + f_2)$, as well as according to the reconstructed digraph, is depicted for a simulation example at three intermediary states.
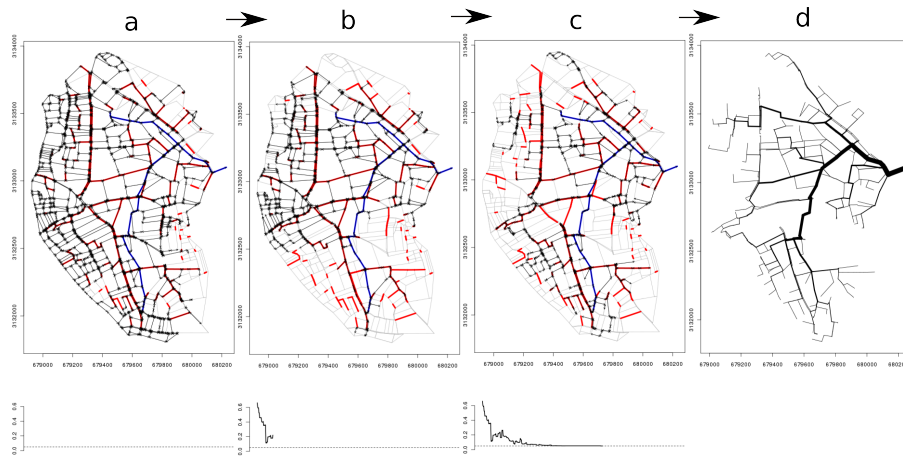
Figure 4: Simulated annealing algorithm 1 intermediary states: a) initial digraph and energy for $k = 0$; b) digraph and energy at iteration $k = 30$; c) digraph and energy at final iteration $k = 309$; d) final obtained ditree. On the top, the lattice edges are depicted in grey; the directed reconstructed graphs are indicated by black arrows. The fixed edges of f with a known direction (© BD-Topo) are depicted in blue, and the fixed edges of f with an unknown direction (remote sensing) are depicted in red. On the bottom, the energy e is depicted as a function of k.

In Figure 4, the cumulative length of the initial digraph d passing through at least 95 % of the f edges covers most of the lattice l and, consequently, has a high energy e due to its overly long cumulative network length. At the intermediary state ($k = 30$), the energy is still high (around 0.2) due to a weak f edge reconnection, and the graph d is segmented in spatial clusters due to a great amount of pruning occurring when the temperature remains high enough. For larger iterations ($k > 200$), the energy slowly decreases with local pruning and branching to converge to the desired maximum level of energy ($e_{max} = 0.05$) at iteration $k = 308$.

4.1.1. Reconstructed network variability

Three forests of 100 reconstructed networks were built using $r, f_2, f_1 + f_2$ to condition the reconstruction. Some of the networks sampled from these three forests are depicted in Figure 5.
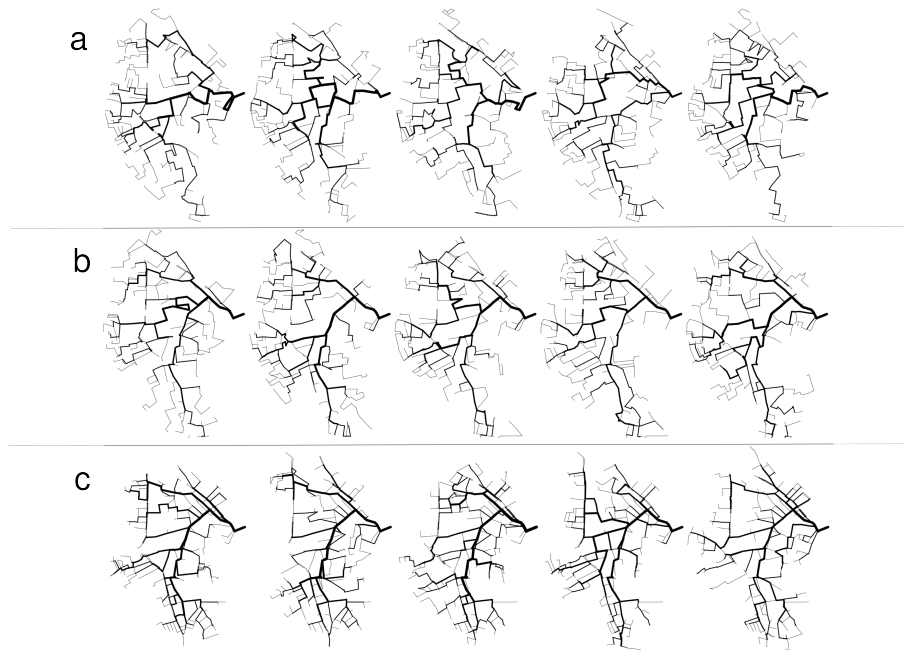

Figure 5: Example of five reconstructed networks using a) r, b) f_2, c) $f_1 + f_2$. All networks are depicted with a line width related to their Strahler orders.

In this Figure, the reconstructed networks look realistic compared to the actual network (Figure 3-b). However, differences appear between cases r, f_2 and $f_1 + f_2$. The topology seems less realistic for r than for $f_1 + f_2$. However, the reconstructed networks look more realistic in the parts of the catchment where the actual main branches belong to the f_2 edges.

4.2. Conditioning data effects on network similarities

Similarities of the reconstructed network forests using the different sets of fixed edges r, f_2 and $f_1 + f_2$ were computed. The h_t, h_c similarity distributions can be deduced from the graphs in Figure 6.
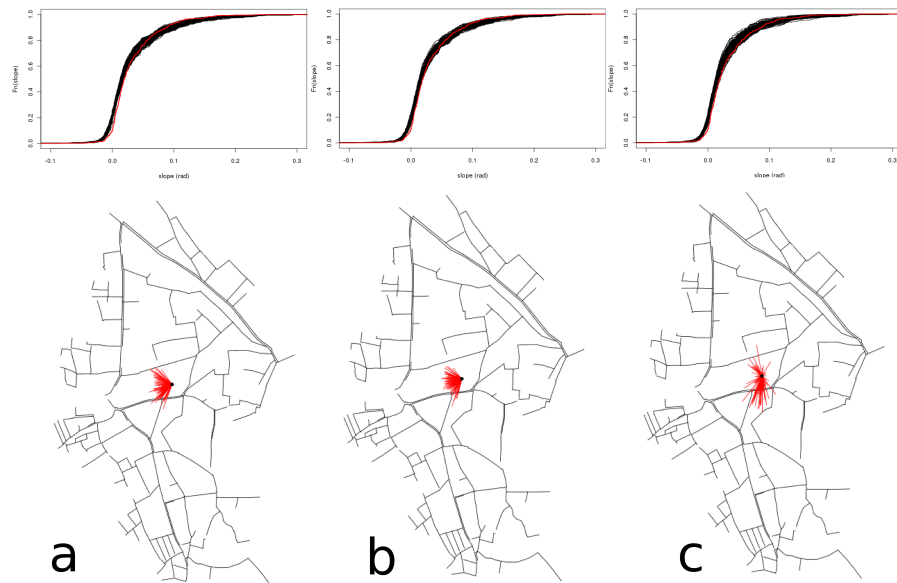

Figure 6: Similarity distribution of forests using a) r, b) f_2, c) $f_1 + f_2$. On the top, the $F_\theta(slope)$ curves are depicted for the actual network (red) and the reconstructed networks (black). On the bottom, the Strahler barycenter of the actual network is depicted by the black point from which red segments join the reconstructed network Strahler barycenters.

The top part of Figure 6 shows that the h_t topographic similarity distribution is quite similar for r and f_2. For $f_1 + f_2$, the slope distribution appears to be of higher magnitude where the slope is greater than 0.2 rad. In all cases, the reconstructed network slopes are not totally centred on the

actual slope distribution and overestimate very low and negative slopes. The bottom part of Figure 6 shows that the h_c topological similarity distribution is once again quite similar for r and f_2, but with a lower magnitude for f_2. For $f_1 + f_2$, the topological dissimilarities appear to be better centred, despite an apparently higher magnitude.

Mean and standard deviation statistics were computed from these different similarity distributions (Table 4.2) and then tested for statistically significant differences, using a two sample comparison Welch T test (Table 4.2). These statistics confirm that the use of f_2 slightly improves the topological and topographical similarities of the networks.

f:	r	f_2	$f_1 + f_2$
Actual edges length (%)	42.6	45.6	57.5
$\bar{L}_d$	23850.5	23868.5	24763
$\bar{h}_t$	0.00101	0.00118	0.00394
σ_{h_t}	0.00201	0.00217	0.00275
$\bar{h}_c$	244.9	230.2	197.5
σ_{h_c}	23.5	22.2	37.5

Table 1: Dissimilarities statistics

Despite having a higher ratio of fixed reaches compared to f_2, the use of $f_1 + f_2$ significantly improves the topological similarity of the networks but a significantly higher topographic dissimilarity also arises (Table 4.2).

23

Probability value H_0:	$r\ vs\ f_2$	$r\ vs\ f_1f_2$	$f_2\ vs\ f_1f_2$
Topographical metric h_t	3.05e-02	8.96e-44	5.91e-38
Topological metric h_c	2.36e-03	3.03e-15	6.33e-10

Table 2: Welch two sample T test probability values

This is probably due to the locations of the f_1 edges corresponding to false remote sensing detection (Bailly et al., 2008), with more being located in the terraced hillslides of the Muscat catchment and favouring higher slopes in the networks. This could also be produced from the necessary e_{max} difference (relaxation), favouring added edges upstream on terraced hillslides.

From the reconstructed forests and similarity distributions, a median network corresponding the reconstructed network with minimized median orders on similarity metrics was computed. This median network is showed using a) r, b) f_2, c) $f_1 + f_2$ on the left of figure 7. In addition, uncertainties in network reconstruction was computed through the frequency in reconstructed networks for each one of the cultivated plot boundaries lattice edge. These uncertainties are also mapped using a) r, b) f_2, c) $f_1 + f_2$ on the right on figure 7. On this figure, the median tree using $f_1 + f_2$ looks the more realistic and close to the actual one. As expected, uncertainties are higher on upstream areas corresponding to terraced hillslides.

5. Conclusions

In this study, a simulated spatial annealing algorithm for the stochastic reconstruction of artificial drainage networks was proposed. This algorithm uses the plot boundary lattice of a cultivated landscape directed by DTM
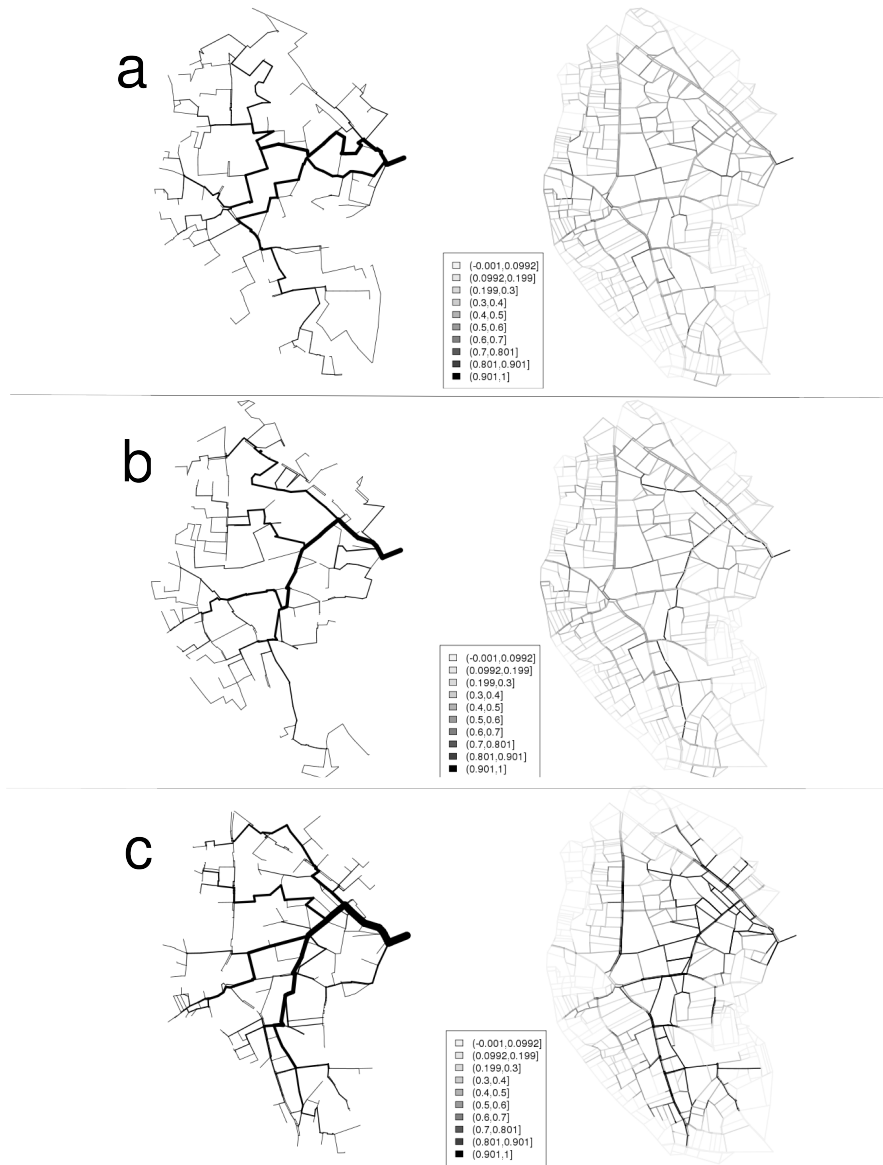
24

Figure 7: The median reconstructed network (left) and frequencies of lattice edges in reconstructed networks (right) using a) r, b) f_2, c) $f_1 + f_2$.

elevation and is conditioned by unconnected known reaches. A number of equi-probable drainage networks connected to a given outlet with an

upstream-downstream ditree structure were produced. They follow a given cumulative length and maximise the reconnection of unconnected known reaches. Randomness in the algorithm comes from upstream-downstream random walks along the cultivated plot lattice and according to random pruning or branching processes in the network.

This algorithm can be used to complete or interpolate from known regions of networks produced from ground observations, geographical databases or remote sensing detection. It also permits the representation of uncertainties in drainage network mapping through multiple realisations of the algorithm.

The results of applying the algorithm to the Muscat catchment with or without using edges detected from airborne scanner laser, from the © BD-Topo database, showed quite realistic reconstructed networks in all cases; the simulations are highly constrained by the 3D structure of the plot lattice. However, a quantitative study of the geometric, topographical and topological properties of the reconstructed networks using different criteria than were used as the objectives for simulated annealing resulted in equivocal conclusions. Adding uncertain information from remote sensing of network locations does not always benefit the model, as it increases the biases of some parameters of the network and reduces others, while adding considerable constraints related to convergence and a long duration to the reconstruction process. It appears that the location, rather than the ratio, of the unconnected true reaches is more important. This assumption could be assessed in future studies using the proposed reconstruction algorithm.

Another question regarding the realistic character of the simulations obtained using this algorithm is related to the final usage of the mapped drainage networks. In this case, only descriptive criteria were used, but cri-

teria based on the hydrological functioning of the simulated networks also need to be assessed. In other words, do uncertainties in drainage network maps resulting from the available data and the proposed algorithm have a significant influence when they are propagated into a given hydrological model ?

The proposed algorithm is generic and can be extended to other networks, to other catchments and even using objectives related to other network properties. Its main requirements are a cultivated plot lattice that can be valued with elevations and known outlets. It also requires a value of the cumulative length of the network to be simulated. This drainage density is expected to be predicted spatially in future studies.

Acknowledgement(s)

Aarts, E., Korst, J., vanLaarhoven, P., 2003. Local search in combinatorial optimization. Princeton: Princeton University Press, Ch. Simulated Annealing.

Amini, J., Blais, M. S. J., Lucas, C., Azizi, A., 2002. Automatic road-side extraction from large scale imagemaps. International Journal of Applied Earth Observation and Geoinformation 4, 95–107.

Bailly, J.-S., Lagacherie, P., Millier, C., Puech, C., Kosuth, P., 2008. Agrarian landscapes linear features detection from lidar: application to artificial drainage networks. International Journal of Remote Sensing 29(12), 3489–3508.

Baltsavias, E., Zhang, C., 2005. Automated updating of road databases from aerial images. International Journal of Applied Earth Observation and Geoinformation 6, 199–213.

Barker, S. B., Cumminc, G., Horsfield, K., 1973. Quantitative morphometry of the branching structure of trees. J. theor. Biol. 40, 33–43.

Beven, K., Binley, A., 1992. The future of distributed models: Model calibration and uncertainty prediction. Hydrological processes 6, 279–298.

Beven, K., Freer, J., 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the glue methodology. Journal of Hydrology 249 (1-4), 11–29.

Bouldin, J., Farris, J., Moore, M., Cooper, C., 2004. Vegetative and structural characteristics of agricultural drainages in the mississippi delta landscapes. Environmental Pollution 132 (3), 403–411.

Carpentier, A., Paillisson, J., Mariona, L., Feunteun, E., Baisez, A., Rigaud, C., 2003. Trends of a bitterling (rhodeus sericeus) population in a man-made ditch network. C. R. Biologies 326, 166–173.

Comoretto, L., 2003. Analyse des chemins de l'eau d'un petit bassin rural sur mnt thr. Master's thesis, Universit de Montpellier II.

Croxton, P., Hann, J., Greatorex-Davies, J., Sparks, T., 2005. Linear

hotspots? the floral and butterfly diversity of green lanes. Biological Conservation 121 (4), 579–584.

Dages, C., Voltz, M., Lacas, J., Huttel, O., Negro, S., Louchart, X., 2008. An experimental study of water table recharge by seepage losses from a ditch with intermittent flow. Hydrological Processes 22, 3555–3563.

Day, W., 1985. Optimal algorithms for comparing trees with labeled leaves. Journal of Classification 1 (1).

Descombes, X., van Lieshout, M., Stoica, R., Zerubia, J., August 2001. Parameter estimation by a markov chain monte carlo technique for the candy-model. In: IEEE Workshop on Statistical Signal Processing. papier invit, Singapour.

Duke, G., Kienzle, S., Johnson, D., Byrne, J., 2006. Incorporating ancillary data to refine anthropogenically modified overland flow paths. Hydrological Processes 20, 1827–1843.

Fairfield, J., Leymarie, P., 1991. Correction to ?drainage networks from grid digital elevation models? Water Resources Research 27 (10), 2809.

Ferraro, P., Godin, C., 2000. A distance measure between plant architectures. Ann. For. Sci. 57, 445?461.

Garcia, M., Camarasa, A., 1999. Use of geomorphological units to improve drainage network extraction from a dem : Comparison between automated extraction and photointerpretation methods in the carraixet catchment (valencia, spain). International Journal of Applied Earth Observation and Geoinformation 1 (3-4), 187–195.

Heipke, C., Mayer, H., Wiedemann, C., Jamet, O., 1997. Evaluation of automatic road extraction. International Archives of Photogrammetry and Remote Sensing 23, 151–160.

Ihaka, R., Gentleman, R., 1996. R: A language for data analysis and graphics. Journal of Computational and Graphical Statistics 5 (3), 299–314.

Janey, N., 1992. Modelisation et synthese d'image d'arbres et de bassins fluviaux associant methodes combinatoires et plongement automatique d'arbres et cartes planaires. Ph.D. thesis, Universite de Franche-Comte.

Jaro, M., 1985. Advances in record linking methodology as applied to the 1985 census of tampa florida. Journal of the American Statistical Society 84 (406), 414–420.

Kruskal, J., 1956. On the shortest spanning subtree of a graph and the traveling salesman problem. Proceedings of the American Mathematical Society 6, 48–50.

Kung, H., Luccio, F., Preparata, F., 1975. On finding the maxima of a set of vectors. Journal of the ACM 22 (4), 469–476.

Lacoste, C., Descombes, X., Zerubia, J., 2005. Point processes for unsupervised line network extraction in remote sensing. IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (10), 1568–1579.

Lagacherie, P., Diot, O., Domange, N., Gouy, V., Floure, C., Kao, C., Moussa, R., Robbez-Masson, J., Szleper, V., 2006. An indicator approach for describing the spatial variability of human-made stream network in regard with herbicide pollution in cultivated watersheds. Ecological indicators 6, 265–279.

30

Lagacherie, P., Rabotin, M., Colin, F., Moussa, R., Voltz, M., 2010. Geomhydas: A landscape discretization tool for distributed hydrological modeling of cultivated areas. Computers & Geosciences 36 (8), 1021–1032.

Lantuejoul, C., 2002. Geostatistical Simulation: Models and Algorithms. Springer Verlag.

Mariani, R., Deseilligny, M. P., Labiche, J., Lecourtier, Y., Mullot, R., 1995. Image Analysis Applications and Computer Graphics. Vol. 1024/1995. Springer Berlin, Ch. Geographic map understanding. Algorithms for hydrographic network reconstruction, pp. 514–515.

Molly, I., Stepinski, T., 2007. Automatic mapping of valley networks on mars. Computers & Geosciences 33, 728–728.

Moslonka-Lefebvre, M., Kervinio, Y., Mejri, R., Pointud, L., 2010. Methodes de comparaison de formes arborescentes: application aux reseaux hydrographiques sur sig. Tech. rep., AgroParisTech-ENGREF.

Moussa, R., Voltz, M., Andrieux, P., 2002. Effects of the spatial organization of agricultural management on the hydrological behaviour of a farmed catchment during flood events. Hydrological Processes 16, 393–412.

O'Callaghan, J., Mark, D., 1984. The extraction of drainage networks from digital elevation data. Computer Vision, Graphics, and Image Processing 28 (3), 323–344.

Pita, R., Mira, A., Beja, P., 2006. Conserving the cabrera vole, microtus cabrerae, in intensively used mediterranean landscapes,. Agriculture, Ecosystems and Environment 115 (3), 1–5.

Prim, R., 1957. Shortest connection networks and some generalizations. Bell System Technical Journal 36, 1389–1401.

Soille, P., Gratin, C., 1994. An efficient algorithm for drainage network extraction on dems. Journal of Visual Communication and Image Representation 5, 181–189.

Spitzer, F., 1970. Principles of the Random Walk. Vol. 34 of Graduate Texts in Mathematics. Springer-Verlag.

Stoica, R., Descombes, X., Zerubia, J., 2004. A gibbs point process for road extraction in remotely sensed images. International Journal of Computer Vision 57 (2), 121–136.

Team, R. D. C., 2009. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.
URL http://www.R-project.org

Thommeret, N., Bailly, J. S., Puech, C., 2010. Extraction of thalweg networks from dtms: application to badlands. Hydrology and Earth Sciences System 14, 1527–1536.

VanGroenigen, J., Stein, A., 1998. Constrained optimization of spatial sampling using continuous simulated annealing. Journal of Environmental Quality 27, 1078–1086.

Varado, N., 2004. Contribution au développement d'une modélisation hydrologique distribuée. application au bassin versant de la donga au bénin. Ph.D. thesis, Université de Grenoble.

Vassilopoulos, A., Evelpidou, N., Bender, O., Krek, A., 2008. Geoinformation Technologies for Geo-Cultural Landscapes: European Perspectives. CRC Press.

Voltz, M., Albergel, J., 2002. Omere: Observatoire mediterraneen de l'environnement rural et de l'eau: Impact des actions anthropiques sur les transferts de masse dans les hydrosystemes mediterraneens ruraux. proposition d'observatoire de recherche en environnement. Tech. rep., Ministere de la Recherche.

Wilson, J., Lam, C., Deng, Y., 2007. Comparison of the performance of flow-routing algorithms used in gis-based hydrologic analysis. Hydrological Processes 21, 1026–1044.

Wright, M., 2010. Automating parameter choice for simulated annealing. Tech. rep., Lancaster university.

Zhang, L., M'Hallah, R., Leung, S., 2010. A simulated annealing with a new neighorhood structure based algorithm for high school timetabling problems. European Journal of Operational Research 203 (3), 550–558.