



**HAL**  
open science

# Méthodes de spatialisation de réseaux arborescents et orientés par simulations conditionnelles : application aux réseaux de fossés de paysages viticoles

Jean-Stéphane Bailly

## ► To cite this version:

Jean-Stéphane Bailly. Méthodes de spatialisation de réseaux arborescents et orientés par simulations conditionnelles : application aux réseaux de fossés de paysages viticoles. Milieux et Changements globaux. Université de Montpellier 2, 2007. Français. NNT: . tel-01430767

**HAL Id: tel-01430767**

**<https://hal.science/tel-01430767>**

Submitted on 10 Jan 2017

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

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



Distributed under a Creative Commons Attribution - NoDerivatives 4.0 International License



Université Montpellier II

Ecole doctorale SIBAGHE  
Spécialité : Terre-Eau-Espace

---

Méthodes de spatialisation de réseaux  
arborescents et orientés par simulations  
conditionnelles : application aux réseaux de  
fossés de paysages viticoles

---

Jean-Stéphane BAILLY

*Thèse de doctorat de l'université,*

*présentée le 18 avril 2007*

**JURY**

Rapporteur :	Patrick DURAND
Rapporteur :	Rachid SENOUSI
Directeur de thèse :	Claude MILLIER
Examineur :	Anne-Véronique AUZET
<i>Invité :</i>	Pascal MONESTIEZ
Examineur :	Philippe LAGACHERIE
Examineur :	Pierre CHEVALLIER



# Table des matières

<b>Introduction Générale</b>	<b>7</b>
<b>1 Contexte et objectifs</b>	<b>11</b>
1.1 Introduction . . . . .	12
1.2 Spatialisation en hydrologie des bassins versants . . . . .	13
1.3 Etudes hydrologiques des réseaux hydrographiques artificiels des paysages cultivés . . . . .	17
1.4 Objectifs d'étude . . . . .	29
<b>2 Hypothèses, démarches méthodologiques et site d'étude</b>	<b>31</b>
2.1 Introduction . . . . .	33
2.2 Hypothèses . . . . .	34
2.3 Démarche méthodologique et objectifs spécifiques . . . . .	44
2.4 Présentation du site d'étude . . . . .	46
<b>3 Simulations de variables à support arborescent et orienté</b>	<b>51</b>
3.1 Introduction . . . . .	53
3.2 Modélisation et simulation géostatistiques de variabilités spatiales le long d'arbres orientés . . . . .	54
3.3 Modélisation et simulation par processus ponctuel de variabilités spatiales par segments homogènes indépendants sur un arbre orienté . . . . .	71
3.4 Bilan . . . . .	86
<b>4 Simulations conditionnelles d'arbres orientés dans un lattice orienté</b>	<b>93</b>
4.1 Introduction . . . . .	95
4.2 Conditionnement : Télédétection de la présence de fossés sur un lattice de limites cadastrales . . . . .	96
4.3 Algorithme de simulation conditionnelle d'un réseau de fossés dans le lattice des limites cadastrales . . . . .	121
4.4 Bilan . . . . .	148
<b>Conclusions et perspectives</b>	<b>151</b>
<b>A Annexes</b>	<b>167</b>
A.1 Annexes du chapitre 3 . . . . .	169
A.2 Annexes du chapitre 4 . . . . .	174



# Remerciements

Ce travail a été réalisé au sein de deux unités de recherche, du CEMAGREF (Institut de recherche pour l'ingénierie de l'agriculture et de l'environnement) et de l'INRA (Institut National de la Recherche Agronomique), dans le cadre, notamment, du projet MOBHYDIC du Programme National de Recherche en Hydrologie.

Il a été financé par l'ENGREF (Ecole Nationale du Génie Rural, des Eaux et des Forêts).

Je suis reconnaissant à Rachid SENOSSI, directeur de recherche en biométrie de l'INRA, et Patrick DURAND, chargé de recherche en hydrologie de l'INRA d'avoir apporté leur point de vue sur cette thèse en acceptant d'en être les rapporteurs.

Je tiens à remercier Claude MILLIER, directeur scientifique de l'ENGREF et de l'INA-PG, maintenant AGRO-PARISTECH, pour avoir suivi régulièrement l'avancement de cette thèse, avoir participé à la phase de validation des résultats, et ce, malgré un éloignement et un emploi du temps peu favorables. Je souhaite exprimer ma gratitude à Philippe LAGACHERIE et Pascal MONESTIEZ pour m'avoir encadré avec une grande disponibilité et clairvoyance tout au long de cette thèse. Je remercie Christian PUECH, Directeur de recherche du CEMAGREF, pour m'avoir également continuellement encouragé. Je remercie également Anne-Véronique AUZET, Professeur des universités de Strasbourg, ainsi que Pierre CHEVALLIER, Directeur de recherche à l'IRD, pour avoir accepté d'examiner ce travail.

Un grand merci Philippe, pour m'avoir accueilli chaleureusement, conseillé et guidé régulièrement mais également pour tes encouragements à publier et clore ce travail. Merci Pascal pour tes conseils avisés et tes bonnes idées.

Merci enfin aux responsables, Marc VOLTZ, Sylvain LABBÉ, Pascal KOSUTH, au personnel, aux thésards, stagiaires de ces deux unités pour leur soutien sans faille, leur assistance technique, administrative et quelquefois sportive. Je pense particulièrement à André MIRALLES, mon co-locataire du bureau 124 de la maison de la télédétection, pour sa précieuse contribution dans le développement d'outils ad-hoc de gestion de données LiDAR et avec qui j'ai pu partager mes angoisses de thésard d'âge mûr.

Finalement, ce travail n'aurait jamais pu se dérouler dans de si bonnes conditions sans le soutien tout en nuances de mes proches : mon épouse et mes filles.

Un peu de simulation conditionnelle pour terminer. A deux des responsables de mon itinéraire et donc de ce travail :  $\omega$ , le hasard, et mon grand-père paternel.



# Introduction Générale

Lorsque l'on s'interroge sur les moyens à mettre en oeuvre pour limiter les inondations ou les pollutions d'origine agricole, la place des aménagements linéaires des espaces cultivés tels les fossés, haies, talus et chemins apparaît souvent comme centrale mais ambiguë. Cette ambiguïté est particulièrement marquée pour les réseaux hydrographiques artificiels créés par les agriculteurs, les services de l'état (génie rural, services de l'équipement) ou les collectivités locales : fossés, chemins, routes, buses enterrées. . . Ces réseaux sont parfois reconnus responsables d'inondations, par effet de concentration des écoulements et d'accélération de la circulation des eaux. Ils sont quelquefois, au contraire, des vecteurs qui favorisent les échanges entre la surface et le souterrain et contribuent ainsi au ralentissement de la dynamique de la crue par rediffusion des écoulements concentrés. Du point de vue de la qualité des eaux, cette ambiguïté est encore plus marquée : tantôt ils piègent ou rediffusent par infiltration les éléments polluants, tantôt ils accélèrent leurs transferts entre leur lieu d'application et le milieu récepteur (une rivière, un fleuve, ou un aquifère).

Les modifications du régime et de la qualité des eaux des hydrosystèmes cultivés liées à l'agencement spatial de ces aménagements et de leurs caractéristiques sont complexes. Cette complexité a de multiples origines. Une première origine est liée aux diverses règles d'implantation, de dimensionnement, d'entretien de ces aménagements qui se traduisent par une forte hétérogénéité spatiale de leurs caractéristiques. Une seconde origine réside dans la diversité de rôles fonctionnels des réseaux artificiels pour différents types de crues, différentes matières actives, différents types de milieux. Les variabilités spatiales et temporelles des processus hydrologiques mis en jeu sur ces réseaux, de leurs singularités qui peuvent parfois à elles seules conditionner leur comportement global, limitent, par non linéarité des réponses et comportements, la généralisation d'études d'impact de ces aménagements sur les hydrosystèmes.

Bien sûr, les recherches sur ces impacts existent, mais elles restent peu nombreuses au plan international. Elles sont souvent locales, expérimentales, et décrivent à l'échelle de l'aménagement les processus élémentaires encore peu connus, notamment sur la qualité des eaux. Ces recherches sont de fait très rarement intégratives à des échelles spatiales de gestion, à des échelles de territoires. Plusieurs causes liées à la complexité de ces hydrosystèmes aménagés limitent cette intégration spatiale, notamment les difficultés d'expérimentation (coût, métrologie) tant du point de vue 1—des mesures hydrologiques, que du point de vue 2—de la description de ces réseaux, dans l'espace. Le besoin de cartographie numérique de ces réseaux apparaît donc important et le manque de méthodes de spatialisation de leur hétérogénéité est un des obstacles majeurs à la généralisation de telles études.



L'objectif général de ce travail de recherche est de proposer des méthodologies qui permettent de représenter spatialement des réseaux hydrographiques anthropiques, de représenter l'hétérogénéité de leurs propriétés locales, à partir de rares données échantillonnées sur le terrain.

Ce travail de thèse s'est déroulé au sein de deux unités de recherche : l'UMR TETIS<sup>1</sup> dont les compétences portent, entre autres, sur les problèmes de spatialisation mettant en oeuvre une démarche métrologique par télédétection, et l'UMR LISAH<sup>2</sup> dont les compétences portent à la fois sur les problèmes de spatialisation de variables dans les paysages ruraux à partir de données rares et sur l'étude hydrologique des réseaux hydrographiques artificiels des paysages agricoles. Ce travail est donc le point de convergence entre ces compétences.

Dans un souci de quantification et de représentation spatiale des incertitudes, l'approche de spatialisation retenue est la simulation spatiale stochastique, i.e. aléatoire. Une fois développée pour les réseaux hydrographiques artificiels des paysages cultivés, cette approche permettrait de mimer les vrais réseaux par des réseaux virtuels réalistes. Par construction, il serait ainsi possible 1– d'intégrer la rareté des données descriptives sur ces réseaux et 2– de représenter les incertitudes de spatialisation. L'avantage principal de réseaux virtuels serait d'être facile à manipuler dans l'espace. Ils pourraient être utiles pour remplacer les données manquantes, tester des protocoles expérimentaux, des plans d'échantillonnage, comparer des scénarios d'aménagement de l'espace, faire des tests statistiques, notamment des études de sensibilité de modèles sur un critère de structure spatiale de réseaux. Quand elles existent, des données complémentaires et localisées, réalisées par exemple par télédétection, pourraient également être introduites dans les procédures de simulation. Ces données pourraient ainsi contraindre les simulations pour en maximiser le réalisme. Dans ce cas, les simulations seraient dites conditionnelles ou conditionnées par ces données.

Si des méthodes de simulation spatiale conditionnelle existent pour des variabilités naturelles et définies dans un plan, en revanche, comment faire lorsque l'espace étudié est linéaire, en réseau, orienté et, qui plus est, artificiel ? Les méthodes définies pour des variabilités naturelles dans un plan sont-elles simplement transférables sur des réseaux ? Comment tenir compte d'un caractère anthropique des variabilités ? La télédétection peut-elle être utile pour contraindre les simulations spatiales d'objets linéaires des paysages cultivés ? Comment juger du réalisme de simulations spatiales de réseaux hydrographiques ? Et suivant quels critères ? Ces questions constituent une déclinaison de l'objectif général dans lequel s'inscrit ce mémoire de thèse.

Ce mémoire s'articule suivant quatre chapitres, qui traitent successivement : (1) d'une revue des méthodes de spatialisation généralement mobilisées en hydrologie puis des études hydrologiques existantes sur l'impact des réseaux hydrographiques artificiels des paysages agricoles ; nous préciserons alors les questions de recherche particulières qui se posent pour la spatialisation de tels réseaux hydrographiques ; (2) d'une présentation des hypothèses et de la démarche méthodologique retenues

---

<sup>1</sup>Territoires, Environnement, Télédétection et Information Spatiale commune à l'ENGREF, au Cemagref et au CIRAD.

<sup>2</sup>Laboratoire d'étude des Interactions Sol-Agrosystème-Hydrosystème commune à l'INRA, l'AgroM et l'IRD.

pour répondre aux objectifs fixés ; nous présenterons également dans ce chapitre les caractéristiques principales du site test retenu ; (3) de développements de méthodes s'appuyant sur des modèles probabilistes spatiaux qui permettent de simuler des variables portées par un réseau hydrographique vu comme un graphe arborescent et orienté ; Ces développements se limiteront à l'étude d'une variable du réseau, sa largeur, prise comme exemple ; (4) de développements de méthodes qui permettent de simuler l'architecture d'un réseau hydrographique dans l'espace géographique en conditionnant ces simulations par des données de télédétection. Ces données permettent de prélocaliser une partie de ces réseaux et d'orienter la simulation.

Ces deux derniers chapitres constituent le cœur méthodologique de la thèse. Ils se concluent chacun par un bilan des méthodes proposées. Des conclusions plus générales ainsi que des perspectives d'utilisation de ces méthodes sont situées en fin de document.



France, 1998  
Pays-Bas, 2002  
Pologne, 2003  
Norvège, 2000

MI 023

FIG. 1 – Arnaud Lesage. *Mutations itinérantes*. Grand prix européen FNAC de la photo, 2005

# Chapitre 1

## Contexte et objectifs

### Sommaire

---

<b>1.1</b>	<b>Introduction</b>	<b>12</b>
<b>1.2</b>	<b>Spatialisation en hydrologie des bassins versants</b>	<b>13</b>
1.2.1	Questions posées à l'hydrologie des bassins versants	13
1.2.2	Les démarches de spatialisation en hydrologie	15
<b>1.3</b>	<b>Etudes hydrologiques des réseaux hydrographiques artificiels des paysages cultivés</b>	<b>17</b>
1.3.1	Réseaux hydrographiques artificiels	17
1.3.1.1	Définitions	17
1.3.1.2	Origines	17
1.3.1.3	Une diversité de fonctions	18
1.3.1.4	Localisation	19
1.3.1.5	Variabilités spatiales	20
1.3.2	Impacts des réseaux hydrographiques sur le régime et la qualité des eaux	20
1.3.2.1	Processus élémentaires	20
1.3.2.1.1	Collecte et distribution des eaux et matières associées	21
	Échanges subsurface-surface	21
	Collectes de surface latérales	22
1.3.2.1.2	Transport et stockage de l'eau : règles hydrauliques	22
1.3.2.1.3	Rétention et transformation des polluants	23
1.3.2.2	Impacts à l'échelle de petits bassins versants	23
1.3.2.2.1	Collecte de l'eau	24
1.3.2.2.2	Transport de l'eau	24
1.3.2.2.3	Dissipation des polluants	26
1.3.2.3	Prise en compte des réseaux artificiels des paysages cultivés dans les modèles hydrologiques actuels	26
1.3.2.4	En conclusion	27
<b>1.4</b>	<b>Objectifs d'étude</b>	<b>29</b>

---

## 1.1 Introduction

---

*Dans ce chapitre, nous introduisons le contexte du sujet de thèse qui est celui de l'hydrologie des bassins versants cultivés.*

*Dans un premier temps, nous regardons dans quel cadre et comment sont abordés les problèmes d'hétérogénéités spatiales internes d'un bassin versant par les hydrologues. Nous nous intéresserons ensuite aux méthodes et outils qu'ils utilisent généralement pour échantillonner, représenter et quantifier ces hétérogénéités, notamment celles de réseaux hydrographiques.*

*Dans un deuxième temps, nous présentons pourquoi et comment les hydrologues étudient les réseaux hydrographiques artificiels des paysages cultivés. Nous extrairons de la littérature les principales caractéristiques de ces réseaux variables dans l'espace et conditionnant le régime et la qualité des eaux 1-à l'échelle de l'aménagement, support de l'acte technique élémentaire, puis 2- à l'échelle de bassins versants.*

*Enfin, en synthèse de ces deux premières sections, nous identifierons les questions méthodologiques particulières que les besoins de représentation des hétérogénéités spatiales de réseaux hydrographiques sous-tendent, questions au coeur des objectifs généraux de cette thèse.*

---

## 1.2 Spatialisation en hydrologie des bassins versants

### 1.2.1 Questions posées à l'hydrologie des bassins versants

La modélisation du cycle de l'eau sur un bassin versant a, par nature, une composante spatiale. Elle s'appuie :

- soit sur quelques caractéristiques physiques globales du bassin versant, ne serait-ce que sa surface ou quelquefois des caractéristiques moyennes, des moments "d'ordre 1" de variables internes aux bassins. C'est par exemple, le cas de formules empiriques qui caractérisent le temps de concentration d'un bassin<sup>1</sup>, sa "nervosité", telle la formule de Passini<sup>2</sup> qui exprime ce temps de concentration  $t_c$  en fonction de la pente moyenne du bassin  $I$ , de la longueur du plus grand thalweg  $L$  et de la surface du bassin  $A$ . On parle dans ce cas de modélisation "globale" ;
- soit sur une discrétisation de l'intérieur du bassin versant pour représenter explicitement les variabilités spatiales internes au bassin. On parle dans ce cas de modélisation distribuée (exemples : TOPOG (Vertessy *et al.*, 1993), MHYDAS (Moussa *et al.*, 2002)) ou semi-distribuée lorsque 1–le nombre de mailles de discrétisation est faible (ex : HYDROTEL (Fortin *et al.*, 2001)), ou 2–lorsque le modèle tient compte d'une distribution (histogramme) non spatialisée de variables internes au bassin (ex : histogramme de l'indice topographique de Kirkby dans TOPMODEL (Beven et Kirkby, 1979)).

Sans polémiquer sur le bien-fondé et les performances entre les approches de modélisations hydrologiques globales ou distribuées, on peut dire que ces deux approches sont complémentaires dans le sens où elles sont destinées à répondre à des questions différentes.

L'histoire de la modélisation hydrologique globale de bassins versants, plutôt statistique, s'est forgée sur les notions de projet, de bilan global de la ressource, dans une démarche d'ingénierie.

Les modèles hydrologiques distribués dans l'espace, plus physiques, ont été développés fortement dans les années 1990. Cet essor semble initialement lié au développement des S.I.G<sup>3</sup>, des M.N.T<sup>4</sup> et des progrès météorologiques (Grayson *et al.*, 2002; Beven, 2001). Au delà de cette impulsion technologique, Grayson *et al.* (2002, p. 1313) souligne que c'est aussi la nature des questions posées aux hydrologues qui a contribué à ce développement : avec la prise en compte des risques hydrologiques (inondations, coulées de boues, pollutions) et dans un souci d'optimisation économique, "on ne s'intéresse plus uniquement à la quantité ou qualité de l'eau en un point de l'espace mais surtout à leurs provenances et leurs déterminants : 1– d'où viennent les contaminants ? ; 2– où investir prioritairement sur le bassin versant pour rectifier les problèmes ?

Pour répondre à ces questions, on a besoin de prédictions des impacts hydrologiques (et écologiques) de l'aménagement du bassin versant qui tiennent compte de la variabilité spatiale interne au bassin (au delà du moment d'ordre 1) et qui soient elles-mêmes spatialisées. Autrement dit, on doit aussi tendre vers une distribution dans l'espace des "sorties" de modèles hydrologiques. C'est déjà clairement le cas,

---

<sup>1</sup>temps de parcours de l'eau du sommet du bassin à l'exutoire

<sup>2</sup> $t_c = 0.108 * \frac{(A*L)^{\frac{1}{3}}}{I^{\frac{1}{2}}}$  en système international avec  $A \geq 40km^2$

<sup>3</sup>Système d'Informations Géographiques

<sup>4</sup>Modèle Numérique de Terrain

par exemple, des modèles hydrauliques de plaine d'inondation (distribution spatiale des hauteurs de submersion). Comme autre exemple pionnier, on peut citer la météorologie, pour qui la prédiction distribuée dans l'espace, liée à la prise en compte croissante des risques, est une demande forte.

Pour le développement et la validation de ces approches distribuées, Hornberger et Boyer (1995), dans leur conclusion de la revue I.U.G.G sur la modélisation hydrologique de bassins versants, relèvent que les hydrologues attendent beaucoup plus de progrès venant de la disponibilité de "données" et paramètres<sup>5</sup> spatialisés (pluie, humidité, débit, conductivités hydrauliques, ...) que de nouveaux modèles. Là encore, un parallèle avec la météorologie distribuée dans l'espace peut-être fait puisque les progrès de ce domaine proviennent en effet largement des efforts faits sur les mesures spatialisées (satellites).

Grayson *et al.* (2002), Western *et al.* (2004) soulignent que, plus que l'exhaustivité spatiale, ce sont surtout les structures spatiales de ces variables et paramètres hydrologiques, i.e. les "moments d'ordre 2" (covariances, connectivités), quelquefois les structures "spatio-temporelles" (Gineste *et al.*, 1998), qui permettent de lever les ambiguïtés, c'est à dire les problèmes d'identifiabilité<sup>6</sup> inhérente à la complexification des modèles (Fig. 1.1) ou de signification physique des modèles (Klemes, 1986). Cette estimation de structure spatiale repose soit sur de nombreuses observations dans l'espace (concepts de "LOP" : Lots Of Points, (Grayson *et al.*, 2002, p.1316)), soit sur des données de télédétection ("surrogate data ", (Grayson *et al.*, 2002, p.1317)).

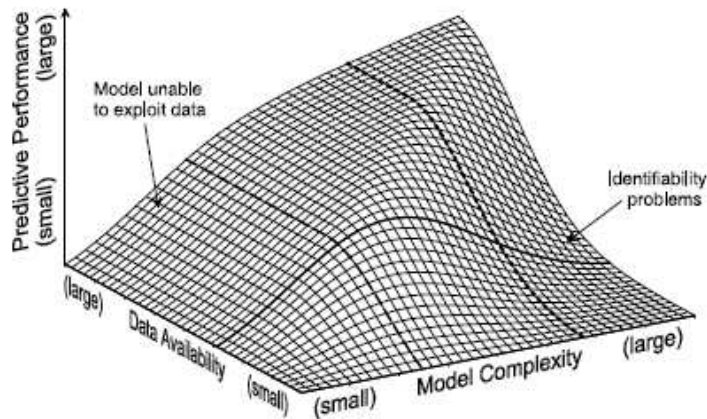


FIG. 1.1 – Relations entre complexité du modèle (pas de discrétisation spatiale), disponibilité de données hydrologiques (spatialisées) et performances de prédiction, d'après Grayson et Blöschl (2000).

<sup>5</sup>au sens inputs de modèles

<sup>6</sup>appelée aussi équifinalité

### 1.2.2 Les démarches de spatialisation en hydrologie

Spatialisation et caractérisation de structures spatiales sont intimement liées. Les approches mises en place pour spatialiser une variable ou un paramètre s'attache en général à faire de l'"estimation" exhaustive, c'est à dire proposer le plus précisément possible une valeur "partout" dans l'espace, en limitant les incertitudes locales de spatialisation.

Classiquement, en hydrologie, on utilise trois approches de spatialisation :

- soit des relevés au sol exhaustifs (données de terrain) correspondant à une approche naturaliste (Ambroise, 1999, p.7) ; on est dans ce cas limité par l'extension spatiale de l'investigation,
- soit par utilisation de moyens de télédétection : les M.N.T pour lier réseaux et surfaces (Charleux-Demargne, 2001), l'occupation du sol à partir de l'imagerie optique, la pluie par radar, l'humidité du sol par radar (Gineste *et al.*, 1998), . . . ; Par la télédétection, on estime la variable d'intérêt au travers de mesures corrélées à cette variable (intensités de réflexion ou d'émission d'ondes électromagnétiques, durées d'aller-retour d'une onde, . . .). On peut dans ce cas être confronté à des problèmes de précision et/ou de résolution spatiale de la mesure.
- soit en s'appuyant sur des schémas de dépendances (relations) spatiales et des échantillonnages partiels dans l'espace et/ou incertains. Ces schémas peuvent être 1-déductifs et se traduire par des schémas d'interpolation spatiale par systèmes experts, logique floue (Cheng *et al.*, 2001) et quelquefois, suivant des fonctions spatiales dites "déterministes" (splines, Thiessen. . .) ou 2-inductifs<sup>7</sup>, souvent géostatistiques. On est, avec ces méthodes, limité à quelques types de variables, distributions, supports.

Souvent, ces outils sont combinés pour améliorer la précision des estimations spatiales (limiter les incertitudes). C'est le cas par exemple lorsque l'on incruste un réseau hydrographique provenant du terrain dans un M.N.T pour améliorer la vraisemblance du cheminement des écoulements de surface d'un bassin (Turcotte *et al.*, 2001). C'est encore le cas lorsque Raclot et Puech (2002) utilisent de l'intelligence artificielle pour améliorer, par contraintes d'écoulements amont-aval, l'estimation de hauteurs d'eau provenant de photographies aériennes sur une plaine d'inondation. Enfin, c'est aussi la philosophie de l'estimation géostatistique par krigeage à "dérive externe"<sup>8</sup> (Chiles et Delfiner, 1999, p.354) lorsque la dérive externe provient d'informations de télédétection, méthode utilisée par exemple par Météo France pour interpoler des champs de pluie à partir de M.N.T (méthode AURELHY (Benichou et LeBreton, 1987)).

Par rapport à ces approches, éventuellement combinées, qui s'attachent surtout à améliorer les estimations locales, Grayson *et al.* (2002, p.1329) notent que, au delà de l'estimation, de l'interpolation, la simulation stochastique de champs de variables et paramètres de modèles hydrologiques qui respecte la structure spatiale d'un champ observé (par ailleurs) est encore plus appropriée car elle permet en plus d'effectuer, au travers de différentes réalisations, des études de sensibilité des modèles hydrologiques à la structure spatiale de ces variables et paramètres. De plus, comme le note Goovaerts (2001, p.23), il y a des "situations où la simulation de champs spatiaux serait bien plus appropriée à la représentation des incertitudes de spatialisation, no-

---

<sup>7</sup>schémas souvent construits par apprentissage sur des hétérogénéités spatiales observées "finement"

<sup>8</sup>la dérive externe est une variable connue partout dans l'espace qui exprime les grandes tendances de variabilité du champs étudié



tamment lorsqu'il y a des transferts latéraux, des flux (par exemple hydrauliques) sur ce champ". Western *et al.* (2001) soulignent que ces méthodes stochastiques, géostatistiques notamment, ne permettent pas encore de capturer, modéliser, représenter des caractéristiques très influentes sur les réponses hydrologiques : celles relatives aux flux latéraux, i.e. aux cheminements des eaux ou réseaux d'écoulements.

En résumé, pour la spatialisation de caractéristiques liées aux réseaux hydrographiques sur un bassin versant, ce sont surtout les approches d'estimation spatiale non stochastiques et/ou par télédétection, à petite échelle (grandes étendues) qui ont été développées. Les démarches de spatialisation par simulation stochastique spatiale semblent d'un intérêt fort pour la modélisation hydrologique distribuée : elles permettent notamment d'effectuer des études de sensibilité des modèles hydrologiques relativement 1- à la structure spatiale des variables et paramètres des réseaux hydrographiques, 2- aux incertitudes de spatialisation de ces réseaux.

Cependant, il n'existe que peu de méthodes de ce type pour les réseaux d'écoulement d'une manière générale, et donc pour les réseaux d'écoulements de surface en particulier.

Avant de spécifier ces besoins méthodologiques, nous allons préciser quelles sont les caractéristiques de ces réseaux étudiées en hydrologie des paysages cultivés, lesquels réseaux correspondent alors à des réseaux artificiels.

## 1.3 Etudes hydrologiques des réseaux hydrographiques artificiels des paysages cultivés

Les réseaux qui supportent les écoulements de surface sur un bassin versant, sont parfois dénommés "chemins des eaux" de surface (Ambroise, 1999, p.10). Ils sont, sur les paysages cultivés, fortement artificiels.

### 1.3.1 Réseaux hydrographiques artificiels

#### 1.3.1.1 Définitions

Il n'existe pas de définition stabilisée du terme réseau hydrographique (absent du dictionnaire). On peut tout de même retenir deux définitions : 1—"On appelle réseau hydrographique, l'ensemble des rivières et autres cours d'eau permanents ou temporaires, ainsi que des lacs et des réservoirs, dans une région donnée" (Glossaire International d'Hydrologie, 1992) et 2—"le réseau hydrographique correspond à tout écoulement linéaire concentré sur la surface topographique qu'il transforme progressivement" (Deffontaines, 1990). L'étude des réseaux hydrographiques se limite donc aux cheminements des eaux de surface concentrées. On étudiera donc par la suite les réseaux de "surface" sensu stricto. Les chemins de sub-surface, artificiels et en amont (drains enterrés), seront très peu considérés.

On entend par réseaux artificiels, des réseaux créés ou induits par la main de l'homme. Dans les paysages cultivés, ils sont constitués de fossés agricoles de bords de parcelle (*ballats*<sup>9</sup> des paysages viticoles Languedociens (Faucon, 1994), *roubines* provençales, . . .), de fossés de bords de route, inter-connectés entre eux par l'intermédiaire d'éléments de voiries et canalisations enterrées au droit des routes et passage d'engins agricoles (Fig. 1.2). Dans les zones de relief et plutôt en tête de bassin, une partie de ces réseaux peut également contenir des petites ravines "naturelles" (Ludwig *et al.*, 1995). Une caractéristique de ces réseaux par distinction aux réseaux naturels est d'être plus intermittents, i.e. en eau et actifs une partie de l'année seulement, car souvent en amont des réseaux naturels.

En France, on peut estimer leur linéaire total à environ 1 200 000 km<sup>10</sup> (contre 290 000 km de fleuves et rivières, source : <http://www.senat.fr/rap/104-271/104-27114.html>).

#### 1.3.1.2 Origines

La genèse de ces réseaux est liée à l'histoire de l'aménagement des espaces cultivés, au gré des réformes agricoles et aménagements collectifs, agricoles ou non (Temple-Boyer, 2000). Typiquement, ils sont le résultat combiné d'un aménagement ancien (antique) qui a façonné la morphologie du paysage actuel (talus, terrasses, voiries) (Berger, 2000), d'opérations de remembrement agricole plus récentes, puis de changements individuels d'utilisation de la sole agricole sur des îlots parcellaires.

---

<sup>9</sup>Fossé bordant les vignes en langue d'Oc

<sup>10</sup>Ordre de grandeur estimé à partir 1—des surfaces drainées citées dans Ollier et Poirée (1981, p. 149) : 200000 ha d'assainissement superficiel et 100 000 ha de drainage enterrés et 2— de la superficie du vignoble français (850 000 ha) et en prenant une moyenne de 100 m de fossé par ha



FIG. 1.2 – Photographies d'éléments de réseaux hydrographiques artificiels : 1-fossé agricole de drainage en pied de talus, 2- fossé de transport(1), 3-fossé de transport(2), 4-élément de voirie, 5-canalisation enterrée, 6-fossé de bord de route, 7-fossé de transport sur un coteau, 8-fossé de bord de route et cannes de provence, 9-fossé-drain principal du réseau de Roujan (34), fortement encombré de végétation.

### 1.3.1.3 Une diversité de fonctions

L'aménagement des espaces ruraux par ces réseaux artificiels est gouvernée par deux grands types de fonctions, dites de "collecte", liées initialement aux flux d'eau (Kao *et al.*, 2002) :

- une fonction de drainage de l'excès d'eau dans les sols cultivés (hydromorphie), associés ou non à des drains enterrés (Carluer et de Marsily, 2005; Nédelec,

2005),

- une fonction de collecte du ruissellement par nappe des eaux de surface, avant que celui-ci ne se concentre et divague sur les parcelles agricoles (Ludwig *et al.*, 1995).

On parle, pour ces deux fonctions "d'assainissement agricole" (Ollier et Poirée, 1981). De ces fonctions de collecte, en découle une troisième : le transport ou transfert d'eau (et polluants), d'amont en aval, pour déboucher au niveau d'un exutoire sur des ruisseaux ou rivières naturelles, appelés "émissaires". D'autres fonctions sont actuellement reconnues à ces réseaux : stockage-diffusion d'eau (Marofi *et al.*, 2002b; LeForner, 2001; Oberlin et Poulard, 2002; Dages, 2006), rétention-transformation ou dissipation des produits phyto-sanitaires (Kao *et al.*, 2002; Louchart *et al.*, 2001; Margoum *et al.*, 2001; Colin, 2000). Ces fonctions sont cependant rarement les fonctions qui sont à l'origine de l'aménagement. Elles induisent plutôt des opérations de réhabilitation et d'entretien de ces réseaux.

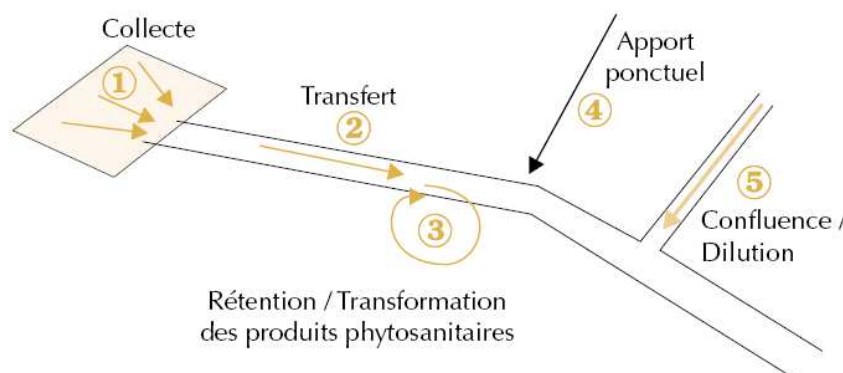


FIG. 1.3 – Fonctions du réseau artificiel de drainage sur la qualité des eaux, extrait de (Kao *et al.*, 2002)

Plus récemment, d'autres fonctions, citées dans la littérature, ont été mises en évidence pour ces éléments des paysages cultivés :

- une fonction écologique par maintien de corridors verts et humides (Pita *et al.*, 2006; Burel et Baudry, 2003; Watson et Ormerod, 2004; Rouquette et Thompson, 2005; Carpentier *et al.*, 2003),
- au travers des échanges induits par ces réseaux entre nappes de sub-surface, ces réseaux artificiels ont également un rôle reconnu dans le bilan hydrique spatialisé des cultures (Guix, 2005).

#### 1.3.1.4 Localisation

Du fait de ces diverses fonctions, on retrouve ces réseaux hydrographiques artificiels principalement dans les zones hydromorphes (zones humides, mouillères, plaine alluviale), où dans des zones où l'énergie du ruissellement de surface peut occasionner des dégâts importants sur les sols : érosion par nappe ou ravinement dans les zones à sols érosifs (Ludwig *et al.*, 1995), érosion par ravinement sur les zones très pentues,

de coteaux généralement en vignoble (Galea et Ramez, 1995). C'est d'ailleurs dans ces grands types de paysages que l'on retrouve les principales études actuelles sur ces réseaux.

#### 1.3.1.5 Variabilités spatiales

Les quelques études qui se sont intéressées à l'hétérogénéité dans l'espace des réseaux hydrographiques artificiels des paysages cultivés ont montré que ces hétérogénéités opèrent à différentes échelles. Elles sont déjà importantes entre différents bassins versants situés dans des contextes différents (Lagacherie *et al.*, 2004). Elles sont importantes sur un bassin du point de vue de la géométrie du réseau et pour des propriétés locales du réseau (Kao *et al.*, 2002). Par exemple, le taux de couverture végétale du réseau, paramètre qui influe sur les transferts hydrologiques comme on le verra ensuite, est très variable et peut passer rapidement de 0 à 90% le long du réseau (Kao *et al.*, 2002; Diot, 2000). De même, des attributs comme la largeur du réseau peut varier dans un rapport de 1 à 10 (Szeleper, 2002; Lagacherie *et al.*, 2004) le long du réseau, même sur de petits domaines (1 km<sup>2</sup>). Pour de plus grands domaines, la "densité" de ces réseaux dans le paysage a été peu étudiée, par manque de données descriptives. D'un point de vue qualitatif, on la sait reliée à plusieurs déterminants, eux-mêmes très variables dans le paysage : les propriétés hydro-dynamiques des sols, le type d'occupation du sol, la morphologie du terrain (Temple-Boyer, 2000).

### 1.3.2 Impacts des réseaux hydrographiques sur le régime et la qualité des eaux

Un premier constat est que le nombre d'études hydrologiques spécifiques sur les réseaux artificiels des paysages cultivés, au plan international, est assez faible. Il est, comparativement, bien développé en France.

#### 1.3.2.1 Processus élémentaires

La structure d'un bassin versant se distingue classiquement suivant deux entités : une surfacique (versants) et une linéaire (réseau hydrographique). Cette "géométrisation" de l'espace du bassin versant est liée en grande partie à deux mécanismes hydrologiques de base : une fonction de production et une fonction de transfert (Galea et Ramez, 1995, p.13). Appliquée aux petits bassins ruraux, on distingue alors les îlots parcellaires (versants) qui portent essentiellement la fonction de production et les réseaux hydrographiques artificiels qui conditionnent la fonction de transfert. Une première nuance doit tout de suite être introduite : du fait de la continuité des échanges entre la matrice environnante du réseau, le sol, et le réseau (Marofi *et al.*, 2002b; Oberlin et Poulard, 2002), le transfert ne doit pas être uniquement vu comme un transport conservatif : toute goutte d'eau arrivant dans le réseau ne sera pas forcément transmise à l'exutoire. Cette fonction de transfert "part" de la collecte des eaux de surface et sub-surface qu'elle stocke, redistribue aux nappes perchées en d'autres points du réseau, ou transporte vers l'exutoire. Du point de vue hydraulique et comparés à d'autres réseaux artificiels (urbains), les réseaux artificiels des paysages cultivés peuvent être vus comme des réseaux qui alternent collecte (assainissement) et distribution.

Pour ce qui est des polluants, les mécanismes de transferts sont les mêmes que pour les flux d'eau lorsqu'ils sont sous forme de solutés dissous, ou ceux du transport

solide lorsqu'ils sont fixés, adsorbés à des particules de sol.

### 1.3.2.1.1 Collecte et distribution des eaux et matières associées

**Échanges subsurface-surface** Carlier et de Marsily (2005) distinguent trois processus élémentaires d'influence des fossés sur les versants pour les échanges d'eau surface-subsurface (Fig. 1.4) : 1–interception du ruissellement hypodermique, 2–rabattement de la nappe par drainage, ou inversement, 3–réinfiltration de l'eau du fossé vers la nappe (Marofi *et al.*, 2002b; LeForner, 2001).

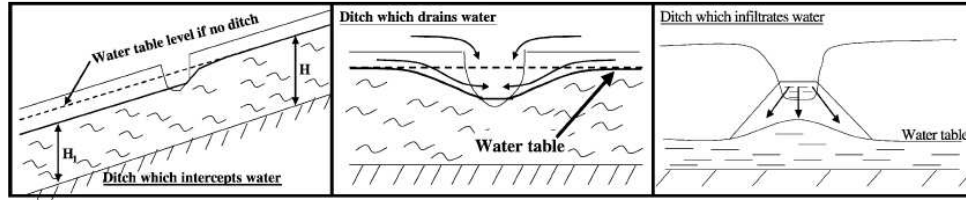


FIG. 1.4 – Influences des fossés sur les versants, extrait de Carlier et de Marsily (2005)

Pour les deux derniers, Marofi *et al.* (2002a) ont montré sur une série d'événements pluviométriques en région méditerranéenne (jusque 120 mm) que les volumes échangés entre fossé et nappe (infiltration puis drainage) peuvent représenter, suivant la position du fossé dans le bassin, 20 à 80 % du volume transitant dans le fossé. Dages (2006) a de même montré que sur des événements pluviométriques et en terme de bilan sur une unité de paysage, 60 % des précipitations sont réinfiltrées vers les nappes par les fossés.

L'impact du réseau sur les échanges d'eau (et matières associées) surface-subsurface est donc très significatif, et ce pour des milieux aux climats contrastés. Cet impact dépend 1– de la position topographique du fossé (fort s'il recoupe la pente, faible s'il est dans le sens de la pente, indifférent si pas de pente) 2–de caractéristiques de la matrice du fossé, i.e. du profil des conductivités hydrauliques à saturation (surtout horizontales) du sol sur la profondeur du fossé et 3– de la profondeur du fossé ou du niveau d'eau libre dans le fossé relativement à celui du toit de la nappe, qui influe sur le niveau de rabattement de la nappe (LeForner, 2001). De même, si le fossé reçoit des collecteurs de drains enterrés, la dynamique des flux à l'interface collecteur-fossé est conditionnée par les caractéristiques de fossé suivantes (Nédelec, 2005, chap.4) : largeur, pente, profondeur (relativement à celle du collecteur), ainsi que de l'azimut du fossé relativement à ceux des collecteurs, ce qui recoupe souvent la direction du fossé par rapport à la pente (les collecteurs sont souvent dans le sens de la pente).

**Collectes de surface latérales** La collecte des eaux (et polluants associés), de surface et "latérales" à un fossé dépend essentiellement de l'occupation du sol et de la topographie des surfaces à proximité et de part et d'autre du fossé.

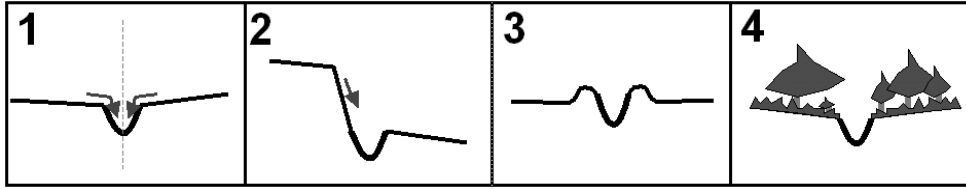


FIG. 1.5 – Situation de collecte (flèches grises) ou non-collecte des eaux et matières associées de surface.

Lagacherie *et al.* (2004) et (Kao *et al.*, 2002) ont résumé cette fonction de collecte latérale suivant la nature de ces surfaces appelées "connexions" latérales. Cette fonction dépend, comme indiqué sur la figure 1.5 :

- **cas 1 et 2** : de l'altitude du fond du fossé par rapport à chaque connexion latérale en rive droite et rive gauche,
- **cas 3** : de la présence de banquettes de part et d'autre du fossé (si présence, pas de collecte d'eau de surface),
- **cas 4** : de la rugosité sur ces surfaces liées à leur degré de végétalisation. Par exemple, dans le cas d'herbicides solubles, si la connection est large, de plus de 6 m, et si elle est, de plus, végétalisée pour plus de 80 % de sa surface, on considère qu'il n'y a pas de collecte latérale possible par le fossé (bandes enherbées) (Madrigal *et al.*, 2002).

**1.3.2.1.2 Transport et stockage de l'eau : règles hydrauliques** Du point de vue hydraulique, la capacité de transport, en régime permanent d'une section de fossé, voirie, canalisation enterrée, à surface libre, décrite par son gabarit, pente, et enherbement, est classiquement évaluée selon la formule de Strickler :

$$Q = K.i^{\frac{1}{2}}.Rh^{\frac{2}{3}}.S \quad (1.1)$$

ou  $Q$  désigne le débit ( $m.s^{-1}$ ),  $i$  la pente du fossé ( $m.m^{-1}$ ),  $K$  le coefficient de Strickler ( $m^{\frac{1}{3}}.s^{-1}$ ), inverse du coefficient de Manning lié à la rugosité,  $Rh = \frac{S}{Pm}$  le rayon hydraulique ( $m$ ), rapport entre la section  $S$  et le périmètre mouillé  $Pm$ .

$Rh$  et  $i$  sont mesurables directement. Le coefficient de Strickler est extrait de tables empiriques ((Galea et Ramez, 1995, p. 28), (Lagacherie *et al.*, 2004, p.271).

Toutefois, cette simple règle présente quelques limites : elle n'est pas applicable aux régimes torrentiels et turbulents (forte pente, nombreuses singularités), lorsque le réseau se met en charge (événements extrêmes dans les canalisation enterrées) (Lesaffre, 1988) et elle n'est pas applicable si l'aval du bief est saturé.

Dans le cas de pentes très faibles (marais, plaines alluviales), la capacité de stockage d'eau dans un bief de réseau dépend de son "volume", i.e. de son gabarit (section maximale) et de sa longueur, aux phénomènes d'infiltration près.

**1.3.2.1.3 Rétention et transformation des polluants** Dans le cas des polluants, on peut là encore se référer aux travaux de synthèse effectués dans le cadre du projet AQUAE (Lagacherie *et al.*, 2004; Kao *et al.*, 2002), axés sur les herbicides. La transformation des polluants peut se faire suivant trois mécanismes : par une voie chimique (hydrolyse, oxydo-réduction), par une voie biologique (actions de micro-organismes) et par une voie photolytique (action de la lumière sur la transformation du polluant) (Kao *et al.*, 2002, p.52). Plus globalement, 3 caractéristiques gouvernent la rétention et la transformation (dégradation) des polluants qui semblent, par ordre hiérarchique d'importance (Margoum *et al.*, 2001) :

- le temps de séjour dans la section de réseau considérée. Ce temps de séjour dépend de la vitesse des écoulements donc de la capacité cinématique de la section  $V = \frac{Q}{S}$ ,
- le taux de matière organique dans le fond du fossé (débris végétaux, herbes) qui agit sur la transformation des polluants,
- le taux de couverture en sédiments dans le fond du fossé qui favorisent les phénomènes d'adsorption.
- le taux d'ensoleillement du fossé qui agit sur la transformation des polluants par photolyse. Ce taux dépend à la fois de l'orientation du fossé par rapport au sud géographique mais aussi de son taux de recouvrement par de la végétation ou d'un relief marqué à proximité.

Pour les polluants adsorbés aux sédiments (référence aux travaux sur l'érosion et le phosphore (Barlow *et al.*, 2003; Ludwig *et al.*, 1995)), ce sont surtout les mécanismes de transport solide, donc la vitesse des écoulements dans le bief relativement à la granulométrie des sédiments qui influe sur leur rétention éventuelle. D'une manière générale, une succession de faibles pentes (série de seuils) favorise le piégeage des sédiments.

### 1.3.2.2 Impacts à l'échelle de petits bassins versants

L'impact des réseaux artificiels sur le régime des eaux d'un bassin versant agit principalement sur la dynamique de la crue. Il se mesure à l'exutoire sur la forme de l'hydrogramme qui caractérise la "nervosité" du bassin suivant des durées caractéristiques, comme le temps de concentration du bassin par exemple, (Fig. 1.6) ou sur le débit maximal.

Pour la qualité de l'eau, l'impact porte aussi sur la dynamique, mais surtout, du fait des dégradations et rétentions, sur les bilans de matières actives exportées du bassin.



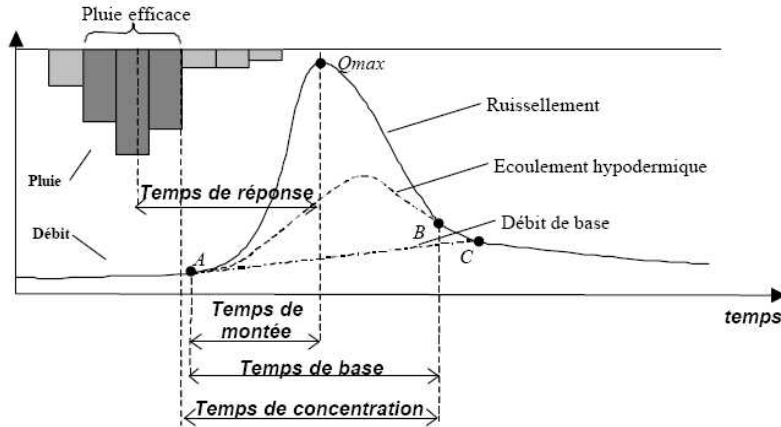


FIG. 1.6 – Hydrogramme et durées caractéristiques, extrait de LeClerc (2004)

**1.3.2.2.1 Collecte de l'eau** Pour l'interception du ruissellement hypodermique et le drainage de la nappe, l'étude de cas menée par Carluer et de Marsily (2005) a montré par simulation sur un petit bassin armoricain à l'aide du modèle ANTHROPOG (dérivé de TOPOG) et pour une petite crue événementielle (60 mm sur 6 jours), que le débit maximal  $Q_{max}$  avec réseau artificiel est 1.6 fois supérieur à celui sans réseau artificiel.

**1.3.2.2.2 Transport de l'eau** On se place ici dans le cadre de l'hydraulique en régime transitoire, c'est à dire de propagation d'une onde de crue.

### *Approches empiriques*

Il existe diverses formules empiriques qui relient la morphologie du bassin et la forme de l'hydrogramme. En premier lieu, on peut citer les formules empiriques sur les temps de concentration d'un bassin (Kirpich, Passini, Ventura...) : toutes sont fonctions de la pente moyenne, de la longueur du plus grand drain et de la surface du bassin. De même, les travaux de Snell et Sivapalan (1994) sur les hydrogrammes unitaires géomorphologiques (G.I.U.H) relient formes de l'hydrogramme et relief du bassin. Dans le cadre des réseaux hydrographiques naturels, il existe une forte corrélation entre relief et forme du réseau hydrographique. Par conséquent, peu d'études portent à proprement parler sur les rapports entre formes de réseaux et hydrogrammes. Toutes se limitent au lien relief-forme de l'hydrogramme.

Pour les réseaux hydrographiques naturels, et dans une approche de régionalisation de modèles globaux Q.D.F (Débit-Durée-Fréquence), Javelle (2001) et Galea et Ramez (1995) ont observés pour un sous-ensemble de bassins inclus dans celui de la Moselle, une bonne relation statistique entre la densité de drainage du réseau hydrographique, rapport entre le linéaire total du réseau hydrographique et la surface du bassin, et un paramètre de "nervosité" du bassin<sup>11</sup>.

<sup>11</sup>la paramètre *Delta* qui décrit la décroissance sur les courbes Q.D.F

Nous n'avons pas trouvé, dans la littérature, d'application de ces méthodes sur les réseaux artificiels des paysages cultivés.

### *Approches physiques : Hydraulique des réseaux hydrographiques*

D'un point de vue hydraulique, le transport de l'eau transitoire dans les réseaux hydrographiques est décrit par le système d'équations de Barré de Saint-Venant unidimensionnelles (conservation de la masse et de la quantité de mouvement). Moyennant quelques hypothèses et suivant le type de réseau, on utilise plus souvent des simplifications de ce système, notamment l'onde cinématique et l'onde diffusive (Moussa et Bocquillon, 1996). Pour les transports d'eau dans les réseaux artificiels des paysages cultivés, on utilise plutôt l'onde diffusive, en y incorporant des flux "latéraux" pour tenir compte des échanges latéraux sur une section de réseau par collecte et ré-infiltration (modèles MHYDAS (Moussa *et al.*, 2002), POWER (Braud *et al.*, 2001)).

Les hypothèses simplificatrices à la base de ces deux dérivées des équations de Barré de Saint-Venant sont les suivantes.

- Dans le cas de l'onde cinématique, la plus simple, relative à la conservation de la masse, les mouvements des eaux sont gouvernés par la friction et l'inclinaison du lit (gravité) ; cette hypothèse néglige l'inertie et le gradient de pression causé par la variation de la profondeur de la crue au long du parcours. Les effets d'influence aval ne sont pas pris en compte (une seule onde, se propageant dans un seul sens) ; elle est adaptée à la simulation du ruissellement supposé se produire en lame mince sur des topographies à forte pente.
- L'onde diffusive tient compte en plus de la pression et elle néglige l'inertie. Les effets d'influence aval sont pris en compte par le biais de la diffusion ; elle est adaptée à la simulation de la propagation d'hydrogrammes longs en rivière à pente faible. Sa formulation générale est, sur un bief  $1D$  indexé par  $x$  :

$$\frac{\delta Q}{\delta t} + C \frac{\delta Q}{\delta x} - q - D \frac{\delta^2 Q}{\delta x^2} = 0$$

avec et deux paramètres,  $C$  la célérité et  $D$  la diffusivité, fonctions (non linéaires) du débit, pente, rugosité, ...  $q$  représente les flux latéraux,  $Q$  le débit,  $t$  le temps.

De nombreuses solutions analytiques de ces équations différentielles ont été proposées (Moussa et Bocquillon, 1996) afin de permettre un routage des eaux plus aisé "numériquement". Souvent appliquées sur des cas uniques (un seul réseau), peu d'études suivant ces schémas de propagation d'ondes dans le réseau se sont intéressées à la sensibilité de l'hydrogramme de crue, de sa forme, en fonction des caractéristiques géométriques, topologiques, i.e. de la morphologie du réseau hydrographique "naturel". Là encore, il n'y a pas dans la littérature d'étude spécifique qui relie morphologie du réseau et forme de l'hydrogramme dans le cas des réseaux artificiels des paysages cultivés.

Pour les réseaux "naturels", Moussa (submitted) a très récemment étudié pour de grands bassins du sud de la France (700 à 5500  $km^2$ ), quelle était la sensibilité d'un hydrogramme de crue à la topologie et aux caractéristiques hydrauliques du réseau. Pour ce faire, il a propagé des ondes de crue dans différents réseaux en utilisant un hydrogramme unitaire sur les versants (noyau d'Hayami) et une onde diffusive pour "router" ces hydrogrammes dans le réseau vers l'exutoire. En fixant les autres paramètres (pluie, caractéristiques hydrauliques locales du réseau, i.e. célérité

et diffusivité), il a notamment étudié l'influence de la topologie du réseau, i.e. de sa complexité en terme d'architecture (nombre de sources, longueur totale, longueur du drain principal), et montré que cette architecture avait une forte influence sur l'hydrogramme de crue. Il a notamment montré qu'il existait sur ces réseaux une relation entre un paramètre de forme du bassin, la distance moyenne de tout point du bassin à l'exutoire via le réseau, et le temps de réponse  $t_r$  du bassin ou la pointe de l'hydrogramme  $Q_{max}$  (dans un rapport de 1 à 3 sur les bassins étudiés). A topologie fixée, il observe qu'en faisant varier globalement les caractéristiques hydrauliques du réseau (célérité uniquement), la sensibilité de l'hydrogramme est assez faible.

### ***Poids des singularités***

Dans les études précédemment citées, faute de données précises localement en dehors de la pente, les caractéristiques hydrauliques du réseau (rugosité, section, . . . ou célérité, diffusivité suivant le type de modèle de transfert) sont traitées d'une manière "homogène". Par analogie avec des réseaux aménagés de types 'urbains', on sait que les singularités locales du réseau (seuils, passage par des sections enterrées pouvant se mettre en charge, rétrécissements créant des embâcles freinant les écoulements, "virages" du réseau, confluences (Ambroise, 1999, p. 33)) conditionnent fortement les écoulements (Nédelec, 2005). Du point de vue statistique, ces singularités hydrauliques liées aux propriétés locales du réseau sont souvent situées dans les queues de distribution : profondeur réduite pour un seuil dans le réseau, gabarit réduit pour un avaloir, forte rugosité hydraulique dans un bief non entretenu et envahi de végétations. Il convient donc d'obtenir "une meilleure précision sur les distributions statistiques des valeurs extrêmes, toujours dangereuses à extrapoler et à négliger, compte tenu de leur influence souvent majeure sur les cheminements préférentiels et les événements exceptionnels" (Ambroise, 1999, p. 98).

**1.3.2.2.3 Dissipation des polluants** Pour les transferts de matières actives exportées sur un bassin cultivé, on peut faire référence aux travaux de Louchart *et al.* (2001) qui a comparé les flux de surface de deux herbicides (Diuron et Simazine) sur deux années consécutives en sortie de parcelles agricoles et en sortie d'un petit bassin versant du vignoble Languedocien. Il en déduit que la majeure partie de ces deux herbicides exportés des parcelles vers le réseau de fossés (5/6 pour le Diuron, 14/15 pour la Simazine) n'atteint pas l'exutoire du bassin, soit par processus de ré-infiltration depuis les fossés vers les nappes, nappes où ces produits sont alors dégradés ou retenus (adsorbés), soit par rétention-dégradation dans les fossés.

### **1.3.2.3 Prise en compte des réseaux artificiels des paysages cultivés dans les modèles hydrologiques actuels**

En hydrologie des bassins cultivés, l'intégration des réseaux hydrographiques est généralement réalisée dans les modèles suivant trois modalités :

- Le réseau conditionne la discrétisation, en mode vecteur, de l'espace du bassin versant. Son dessin dans l'espace géographique intervient sur la fonction de collecte des eaux et matières associées de subsurface et surface mais il n'y a pas de "routage" de ces eaux et matières collectées dans le réseau (pas d'hydraulique dans le réseau). C'est par exemple le cas du modèle ANTHROPOG (Carlier et de Marsily, 2005), modèle adapté aux transferts de subsurface "lents" vers le réseau, comparés aux transferts dans le réseau.

- Les modèles raster<sup>12</sup> représentent le réseau sous forme de cellules contiguës de transfert à pas régulier (modèle ANSWERS, (Beasley et Huggins, 1981)) ou irrégulier (modèle MC (Girard *et al.*, 1981)) et avec une fonction de transfert d'eau dans ces cellules suivant l'équation de Strickler (Chahinian, 2004, p.7).
- Le réseau et ses propriétés locales (largeur, profondeur...) sont explicitement pris en compte dans le modèle sous forme vecteur, suivant des unités homogènes. Dans ce cas, et pour la partie hydraulique il y a une fonction de transfert dans le réseau, généralement basée sur l'onde diffusante (modèle MHYDAS (Moussa *et al.*, 2002)), ou l'onde cinématique (TOPOG (Vertessy *et al.*, 1993)), plus rarement suivant la formule de Strickler (modèle THALES (Grayson *et al.*, 1992)). Des approches de modélisation basées sur la constitution d'indicateurs spatiaux, par exemple pour le transfert d'herbicides dans le bassin (Lagacherie *et al.*, 2004), peuvent aussi utiliser cette représentation explicite des réseaux et de leurs propriétés suivant des unités homogènes.

En résumé, peu de modèles hydrologiques actuels intègrent explicitement les réseaux hydrographiques artificiels. De plus, c'est principalement l'emplacement du réseau dans l'espace géographique qui est le plus fréquemment pris en compte dans les outils de modélisation hydrologique car il conditionne la segmentation et la topologie de l'espace du bassin versant. Bien que nombre de propriétés locales le long du réseau soient désormais expérimentalement reconnues comme importantes dans le fonctionnement, notamment pour la qualité des eaux, elles sont peu souvent prises en compte dans les modèles. Lorsqu'elles ne sont pas absentes des modèles, ces propriétés peuvent également être traitées globalement, comme paramètres d'ajustement, car souvent considérées comme non accessibles et/ou difficilement mesurables (exemple : conductivité, rugosité hydrauliques).

Des travaux et projets actuels (exemple du projet SEVE) tentent, pour plusieurs échelles, de mieux lier la perception des processus moteurs expérimentalement à la conception de modèles pour des bassins ruraux fortement aménagés, en intégrant plus de processus et donc de variabilités spatiales, notamment sur les réseaux.

Pour les variables de réseaux hydrographiques, on peut prendre ici l'exemple des modèles hydrologiques d'assainissement urbain. Ces modèles (exemple : CANOE), développés eux aussi dans un contexte fortement aménagé, intègrent tous à la fois le dessin du réseau dans l'espace et ses propriétés locales (section, rugosité, pentes, singularités).

#### 1.3.2.4 En conclusion

Aux échelles spatiales étudiées (petits bassins cultivés), beaucoup de caractéristiques des réseaux artificiels influent significativement sur le régime et la qualité des eaux d'un bassin versant. On peut distinguer deux catégories de caractéristiques :

- celles relatives à l'"architecture" du réseau, i.e. sa topologie (distribution des sources et des ramifications qui créent des singularités) et sa géométrie (la longueur du réseau), sa topographie et son "orientation" : la distribution des pentes du réseau ; la distribution des azimuts<sup>13</sup> des segments de réseaux.
- les attributs "locaux" du réseau qui sont conditionnés entre autres par les actes techniques élémentaires des acteurs de ce milieu. Pour les principales, on peut citer : la présence de singularités hydrauliques liées au gabarit de la section en travers du réseau, aux ruptures de pentes ; le caractère rugueux ou lisse du

<sup>12</sup>Segmentation de l'espace sous forme de grille à mailles ou cellules régulières

<sup>13</sup>direction par rapport au nord géographique

réseau lié à l'encombrement du lit par des végétaux, blocs, débris ; les propriétés hydro-dynamiques (conductivités hydrauliques) de la matrice du lit du réseau ; la présence de matière organique dans le réseau (débris végétaux, couverture végétale).

L'impact hydrologique de ces réseaux n'a été évalué, on le voit, que sur des bassins de l'ordre du  $km^2$ . L'expérimentation sur ces impacts (comparaison de situations avec/sans aménagement, avec changements des caractéristiques du réseau) étant difficilement envisageable à ces échelles, ces études s'appuient le plus souvent sur des simulations numériques issues de modélisations distribuées à base physique.

Même s'il faut regarder ces résultats de changement d'échelle avec prudence, car issus de simulation numérique, d'analogies hydrauliques et rarement d'observations in situ, il est intéressant de noter que :

1. En reprenant les idées évoquées par Grayson *et al.* (2002), pour mener à bien ce changement d'échelle dans un modèle hydrologique, il faudrait éprouver le modèle sur un grand nombre de situations, multi-locales, multi-"tailles" de bassins et donc disposer d'un grand nombre d'échantillons, un échantillon étant alors un bassin élémentaire avec son réseau, son parcellaire. . . décrits dans l'espace et accompagnés de mesures hydrologiques.
2. Dans le cas d'une démarche "désagrégative" (Colin, 2000), bien que difficile<sup>14</sup> car il faudrait se placer dans des situations de réseaux aux caractéristiques tranchées "toutes choses étant égales par ailleurs" (pluie, . . .), cet accroissement du nombre d'échantillons dans l'espace serait à fortiori nécessaire.
3. Même pour de petits bassins ( $1 km^2$ ), l'effort de spatialisation, par des relevés terrain des variables et caractéristiques des réseaux (Lagacherie *et al.*, 2004) est important<sup>15</sup>.

---

<sup>14</sup>Sinon, il devient vite difficile dans les signaux hydrologiques de décrypter la part relative à telle ou telle caractéristique variable du bassin

<sup>15</sup>estimé à environ 10 jours-homme / $km^2$

## 1.4 Objectifs d'étude

Les marges de progression de l'évaluation des impacts hydrologiques des réseaux hydrographiques artificiels des paysages cultivés à l'échelle de bassins versants sont intimement liées aux possibilités de spatialisation de ces réseaux. Cette spatialisation doit déjà être facilement reproductible pour des échelles spatiales fines afin de pouvoir représenter les réseaux de multiples bassins élémentaires de quelques  $km^2$ . A terme cette spatialisation devra être extrapolée, suivant des modes certainement différents, vers des bassins de gestion (de l'ordre de la centaine de  $km^2$ ). Pour ces échelles fines, on ne peut raisonnablement envisager cette spatialisation que suivant des méthodes économiquement acceptables et selon une résolution spatiale qui corresponde à la dimension physique du support des actes techniques élémentaires sur ces réseaux (unité d'aménagement, d'entretien), actes qui peuvent générer des singularités de réseaux très influentes sur leur fonctionnement.

Comme nous l'avons vu dans les deux premières sections, il apparaît important de veiller à ce que les méthodes de spatialisation mises en oeuvre respectent les structures spatiales ainsi que les histogrammes des variabilités observées, propriétés globales qui conditionnent le fonctionnement hydrologique de ces réseaux. Pour ce faire, une démarche capable de combiner différentes approches de spatialisation doit être envisagée : approche par télédétection ; approche basée sur un échantillonnage de terrain associé à une modèle de répartition spatiale. Par exemple, une fois évaluée la part observable à moindre coût par télédétection de cette variabilité, une démarche de simulation stochastique du "reste" de la variabilité permettrait de répondre aux critères précédents. Cette démarche, au travers des réalisations multiples des simulations représenterait explicitement les incertitudes de spatialisation et permettrait d'envisager des tests de sensibilité de modèles hydrologiques.

Pour des champs spatiaux structurés en réseaux, cette démarche combinée dite de simulation conditionnelle, suivant différents modes de conditionnement, n'existe pas et reste à proposer.

Développer des méthodes et outils de simulation spatiale conditionnelle pour les réseaux hydrographiques, l'appliquer aux réseaux artificiels des paysages cultivés, constituent les perspectives dans lequel s'inscrit ce travail de thèse.



## Chapitre 2

# Hypothèses, démarches méthodologiques et site d'étude

### Sommaire

---

<b>2.1</b>	<b>Introduction</b>	<b>33</b>
<b>2.2</b>	<b>Hypothèses</b>	<b>34</b>
2.2.1	Un réseau hydrographique artificiel vu comme un arbre orienté	34
2.2.1.1	Un arbre orienté en théorie des graphes	35
2.2.1.2	Vocabulaire sur un arbre orienté, modèle de réseau hydrographique	37
2.2.2	Un réseau hydrographique artificiel, sous ensemble du lattice des limites de parcelles cadastrales	37
2.2.3	Simulations spatiales conditionnelles : définitions et principes	38
2.2.3.1	Simulation et simulation conditionnelle	38
2.2.3.1.1	Simulation spatiale	39
2.2.3.1.2	Simulation spatiale stochastique	39
2.2.3.1.3	Simulation spatiale conditionnelle	40
2.2.3.2	Simulation conditionnelle <i>vs</i> estimation spatiale	41
2.2.3.2.1	Des différences de propriétés locales ou globales	42
2.2.3.2.2	Simulation : représentation plus explicite des incertitudes de spatialisation	43
<b>2.3</b>	<b>Démarche méthodologique et objectifs spécifiques</b>	<b>44</b>
2.3.1	Développer des méthodes de simulations et simulations conditionnelles de variables locales sur un arbre orienté	44
2.3.2	Développer des méthodes de simulations conditionnelles de l'architecture des réseaux	44
2.3.2.1	Explorer les possibilités de détection de ces réseaux artificiels sur des données LiDAR	44
2.3.2.2	Simuler une architecture de réseau réaliste à partir d'une connaissance partielle du réseau dans le lattice cadastral orienté par un M.N.T	44
<b>2.4</b>	<b>Présentation du site d'étude</b>	<b>46</b>



2.4.1	Situation . . . . .	46
2.4.2	Caractéristiques du milieu . . . . .	48
2.4.3	Fonctionnement hydrologique . . . . .	49
2.4.4	Les différents sites tests utilisés . . . . .	49

---

## 2.1 Introduction

---

*Dans ce chapitre, nous introduisons les principales hypothèses méthodologiques qui permettent ensuite de préciser la démarche retenue pour répondre aux objectifs généraux présentés en fin du chapitre 1.*

*Une première hypothèse de travail est de considérer les réseaux hydrographiques artificiels des paysages cultivés comme des graphes arborescents et orientés. Nous considérerons ensuite que ces graphes arborescents constituent un sous ensemble d'un graphe plus général, maillé et représentant les limites des parcelles cadastrales. Nous introduirons dans cette section du vocabulaire sur ces graphes que nous utiliserons continuellement par la suite. Les notations mathématiques liées à ce vocabulaire seront, quant à elles, introduites au fur et à mesure des besoins dans les chapitres suivants.*

*Nous présenterons également dans cette première section la philosophie générale de la démarche d'une simulation spatiale stochastique, conditionnelle ou non conditionnelle, par comparaison à ce qui se fait plus classiquement par estimation spatiale. Enfin, sur la base de ces hypothèses et dans le cadre d'une démarche par simulation spatiale conditionnelle, nous préciserons les objectifs spécifiques de thèse.*

*Dans un troisième temps, et avant d'entamer la mise en oeuvre de cette démarche, nous présenterons dans les grandes lignes les caractéristiques du site d'étude, site largement présenté par de nombreux travaux antérieurs.*

---

## 2.2 Hypothèses

### 2.2.1 Un réseau hydrographique artificiel vu comme un arbre orienté

Les réseaux hydrographiques ont une structure graphique dans l'espace qui s'apparente le plus souvent à une structure hiérarchique ou arborescente, i.e. sans maille ou boucle. Autrement dit, pour de tels réseaux, il n'y a qu'un et un seul chemin qui relie un point du réseau à un autre. Cette propriété graphique n'est pas systématique (divagation dans les deltas, rivières en tresse) mais elle l'est très fréquemment dès lors qu'il y a du relief. Pour les réseaux naturels et pour de grandes étendues, c'est très souvent le cas. La plupart des algorithmes de drainage qui permettent l'extraction de ces réseaux à partir de M.N.T sont d'ailleurs construits sur cette hypothèse.

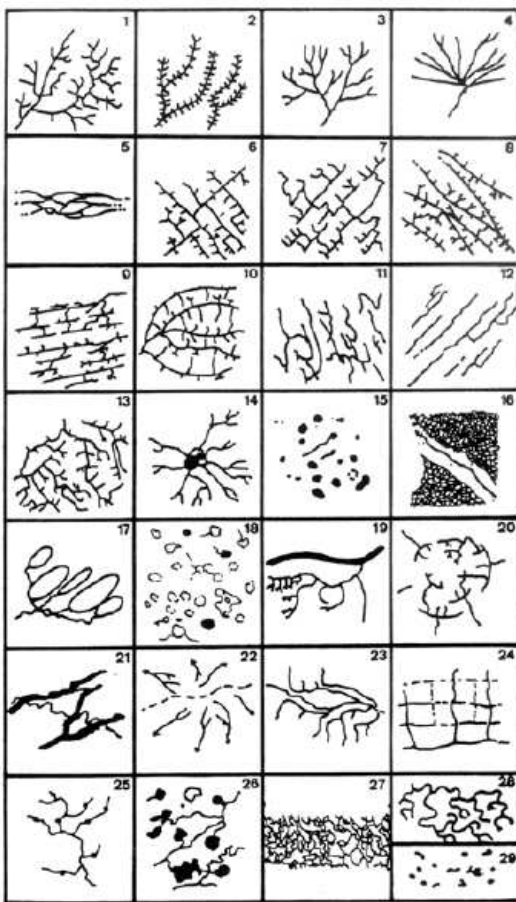


FIG. 2.1 – Typologie de réseaux hydrographiques selon Howard (Howard, 1967) : Les types 6,7,8,9 sont des types de treillis simples, directionnels, de failles, de joints, extrait de (LePape, 1998).

Qu'en est-il des réseaux artificiels ? Pour nombre de ces réseaux, cette propriété n'est pas du tout apparente graphiquement comme le démontre la figure 2.2 (haut-gauche) qui représente le réseau hydrographique artificiel du bassin viticole de Roujan (Lagacherie *et al.*, 2004) dans l'espace géographique. En revanche, elle l'est dès lors que l'on s'intéresse à l'orientation du réseau dirigée par les écoulements. Du point de vue de la classification descriptive des réseaux hydrographiques naturels de Howard (Fig. 2.1), ils s'apparentent au type "treillis", notamment directionnel, marquant les systèmes (naturels) sur roches sédimentaires à pendage net.

On considérera donc qu'un réseau hydrographique artificiel d'une zone cultivée peut être fonctionnellement considéré comme un arbre, une fois l'orientation des segments du réseau introduite. Ce réseau est physiquement orienté ou dirigé par le sens d'écoulement, de l'"amont vers l'aval" (notion guidée par le sens d'écoulement plus qu'une altitude, surtout en zone peu pentue<sup>1</sup>). On parle alors d'arbre orienté. La plupart des modèles hydrologiques qui intègrent ces réseaux les considèrent comme arborescents et orientés, ce qui, par ailleurs, simplifie les équations de l'hydraulique par rapport au cas maillé<sup>2</sup>. La figure 2.2 (haut-droite) représente la réseau de Roujan, avec cette fois-ci, son orientation. Un zoom réalisé sur une partie du réseau montre bien la différence entre une structure apparente maillée dans l'espace géographique et une structure topologique fonctionnelle arborescente, sans maille.

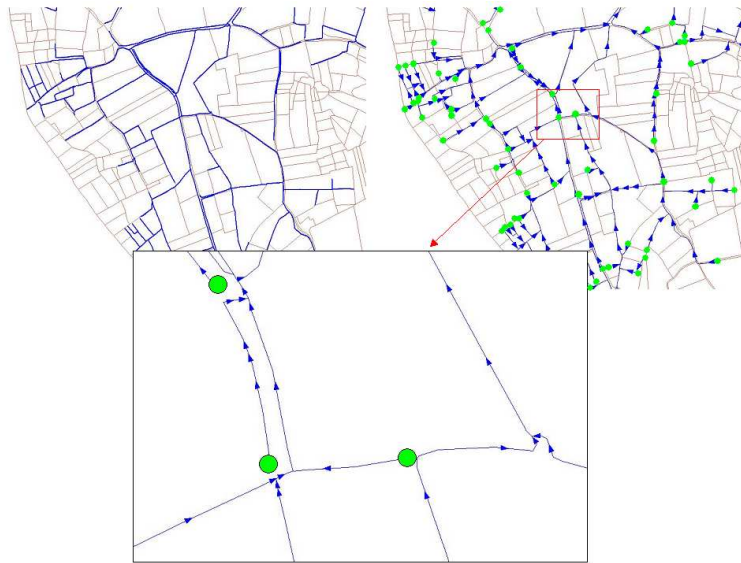


FIG. 2.2 – Dessin dans l'espace géographique du réseau de Roujan avec et sans orientation. Les points en vert clair représentent les sources du réseau

### 2.2.1.1 Un arbre orienté en théorie des graphes

Un graphe est un ensemble de noeuds (sommets) dont certaines paires sont reliées par des arêtes. Un arbre est un graphe connexe acyclique, sans boucle (Fig. 2.3). Un

<sup>1</sup>Les écoulements à contresens de la pente d'un bief sont possibles.

<sup>2</sup>Dans ce cas on doit tenir compte d'une répartition de débit dans la maille qui permette un équilibre de la charge hydraulique dans la maille.

arbre à  $n$  noeuds contient  $n - 1$  arêtes.

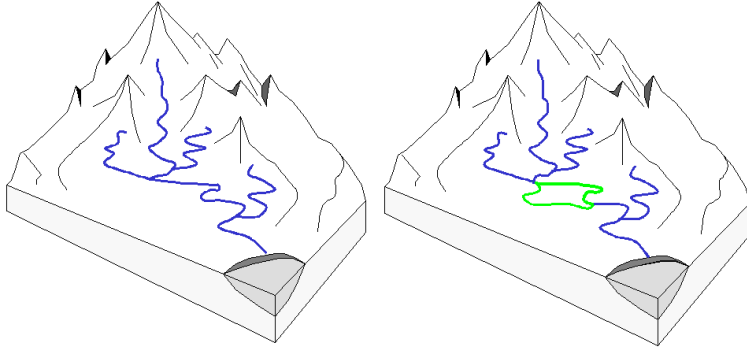


FIG. 2.3 – Exemple de réseau hydrographique arborescent (gauche) et maillé (droite) : la maille est en vert.

Un arbre est orienté si chacune de ses arêtes est un vecteur orienté d'un noeud de départ vers un noeud terminal. On parle dans ce cas d'arborescence (rooted tree). Pour chaque noeud de l'arborescence, on peut calculer son degré  $d$  (Fig. 2.4), i.e. le nombre d'arêtes distinctes qui ont ce noeud comme noeud de départ ou de terminaison. On parle de degré inférieur  $d_-$  (degré sortant), pour le nombre d'arêtes distinctes qui ont ce noeud comme noeud de départ et de degré supérieur  $d_+$  (degré entrant), pour le nombre d'arêtes distinctes qui ont ce noeud comme terminaison. Un arbre orienté construit donc un ordre sur les arêtes, contient un puits (noeud vers lequel pointent tous les autres noeuds du graphe) ou une racine (noeud qui pointe vers tous les autres noeuds du graphe), et des feuilles, noeuds de degré  $d = 1$ . Stricto sensus, on appelle racine, le noeud de degré supérieur nul, et un puits, celui de degré inférieur nul. Les sources sont des feuilles de degré supérieur nul (Fig. 2.4 et 2.5). Compte tenu de l'orientation d'un réseau hydrographique, on parlera arbitrairement par la suite de racine de l'arbre comme d'un puits (degré inférieur nul).

On parle de valuation d'un graphe ou d'un arbre, dès lors que les noeuds ou arêtes du réseau supportent une information sémantique, qualitative ou quantitative. Cette information sémantique portée par un noeud peut-être un attribut du réseau conceptuel ou physique modélisé par l'arbre : le nom de famille de chaque membre d'un arbre généalogique par exemple. La valuation des arêtes peut porter par exemple sur la dimension physique de l'arête (longueur du segment de rivière en km). La valuation de l'arbre peut porter à la fois sur les arêtes et les noeuds, ce qui sera notre cas.

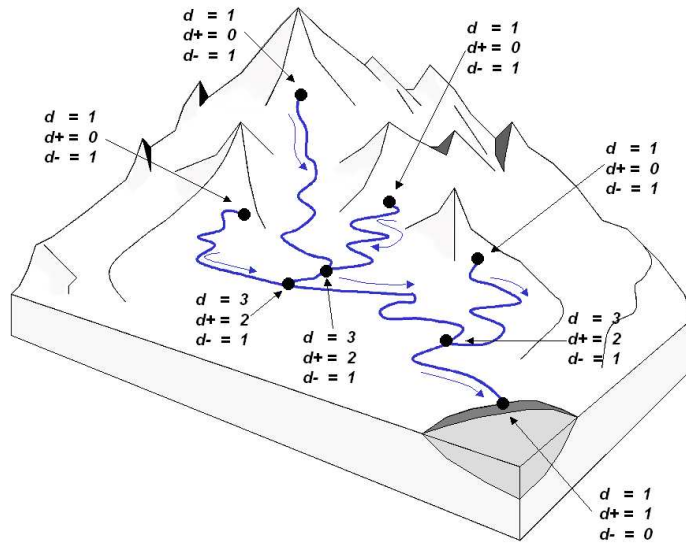


FIG. 2.4 – Calcul des degrés ( $d$ ), degrés inférieurs ( $d_+$ ) et supérieurs ( $d_-$ ) sur les noeuds d'un réseau (points noirs)

### 2.2.1.2 Vocabulaire sur un arbre orienté, modèle de réseau hydrographique

Lorsqu'un arbre représente physiquement un réseau hydrographique, on peut faire une analogie de vocabulaire. Par exemple, on parlera par la suite :

- de l'exutoire (outlet) comme du puits (racine) de l'arbre (Fig. 2.5),
- de confluences (junctions) comme des noeuds de l'arbre de degré supérieur ou égal à 2 (Fig. 2.5) ; on parle quelquefois de noeuds "intérieurs" du réseau,
- de rivières, comme d'une succession d'arêtes amont-aval partant d'une source jusqu'à une confluence où une autre rivière, de plus grande longueur est rencontrée (définition usuelle d'une rivière en géographie) (Fig. 2.5),
- d'un chemin amont-aval comme d'un cheminement au fil de l'eau d'un point quelconque vers un autre point situé plus en aval.
- d'une branche amont d'un point comme de l'ensemble du réseau qui se jette en ce point,
- de rivières parallèles lorsqu'aucun point de l'une n'est situé sur la branche amont de n'importe quel point de l'autre,
- du fleuve ou rivière principale comme de la rivière au plus grand linéaire qui débouche à l'exutoire (Fig. 2.5).

### 2.2.2 Un réseau hydrographique artificiel, sous ensemble du lattice des limites de parcelles cadastrales

Lorsque l'on observe la localisation des réseaux hydrographiques artificiels des espaces cultivés, ils apparaissent clairement et exclusivement situés sur les limites des parcelles cadastrales (plus que culturales) comme le montre la figure 2.6 d'une partie de 8  $km^2$  du bassin versant cultivé de la Peyne. Une explication vient du fait que la morphologie du terrain conditionne conjointement l'emplacement des limites

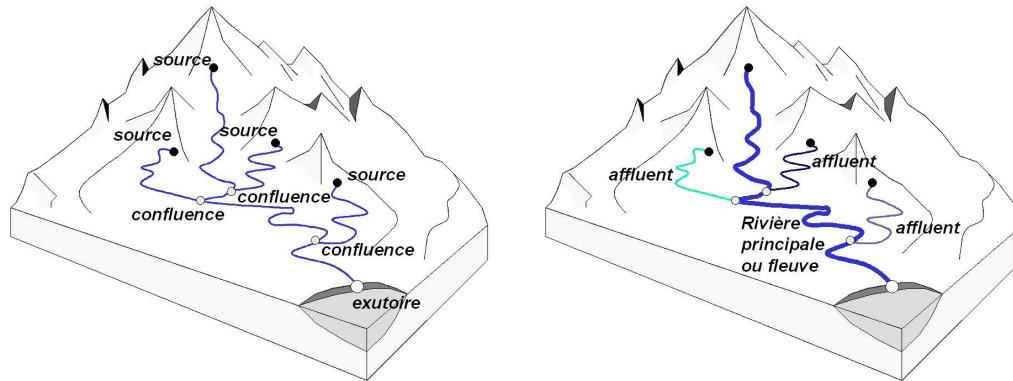


FIG. 2.5 – Exemple de réseau avec situation des sources, confluences et exutoires (gauche) et des différentes rivières suivant un code couleur (droite).

cadastrales et des réseaux d'écoulement. Seuls les passages par canalisations enterrées sous routes et chemins d'exploitation ne correspondent pas à une de ces limites mais à une inter-connection entre îlots parcellaires. En conséquence, lorsqu'on s'intéresse à ces réseaux, on peut focaliser son attention sur ces limites cadastrales et réduire l'espace exploré à ces limites. On considérera que cette information sur l'emplacement des limites cadastrales dans l'espace est donnée, et facilement généralisable. En effet, les cadastres sous-forme numérique tendent à se généraliser et nombre de travaux de spatialisation utilisent désormais ces données (Durupt et Taillandier, 2006; Trias-Sanz et Deseilligny, 2004).

On appelle lattice ou treillis des limites parcellaires le réseau maillé, cyclique, planaire et connexe formé par ces limites. Une arête ou limite parcellaire relie deux noeuds du réseau. Chaque noeud a un degré supérieur ou égal à trois.

### 2.2.3 Simulations spatiales conditionnelles : définitions et principes

Comme il a été mentionné dans le chapitre 1, reproduire une variabilité spatiale "réaliste", notamment en terme de distribution (histogramme), est un critère de spatialisation important. Afin de mieux représenter les queues de distribution (histogramme), la simulation spatiale est un outil préférable à l'estimation. Nous précisons ci-après les nuances entre ces termes et les raisons de cette propriété, en prenant pour exemple, le cadre formel de la simulation et de l'estimation géostatistique.

Dans la suite du manuscrit, sous les termes simulation, (conditionnelle ou non conditionnelle) et estimation, on entendra implicitement simulation et estimation sous forme numérique et spatiale.

#### 2.2.3.1 Simulation et simulation conditionnelle

Par définition, une simulation cherche à mimer la réalité. Lorsque l'on cherche à simuler un système physique réel, on s'appuie alors sur une technique informatique, typiquement un algorithme, qui peut être, par exemple, développé à partir d'une modélisation mathématique de ce système.

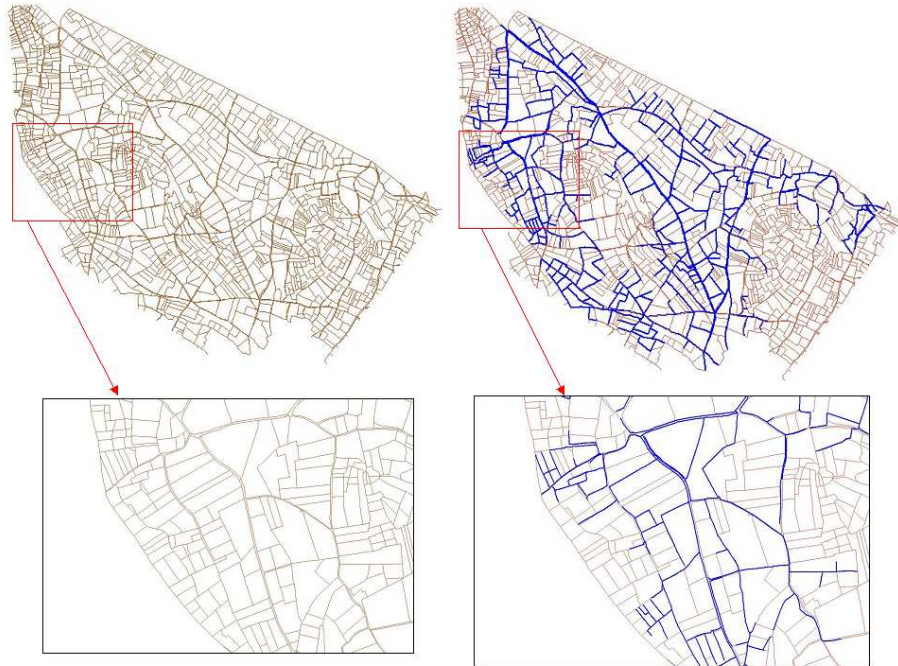


FIG. 2.6 – Haut : Lattice des limites cadastrales (brun) et réseau hydrographique artificiel (bleu) sur le bassin du Bourdic (34) ; Bas : zoom sur une partie du bassin de Roujan.

**2.2.3.1.1 Simulation spatiale** Une simulation spatiale mime un système qui est indexé (géoréférencé) dans l'espace. Elle s'effectue à partir d'un modèle mathématique qui décrit la structure spatiale de ce système. La structure spatiale est ici entendue comme le résultat de processus spatiaux, i.e. de dépendances spatiales qui expriment une variabilité spatiale. Ces modèles peuvent être déterministes, sans caractère aléatoire. C'est le cas par exemple de nombre de modèles hydrauliques mais aussi des modèles utilisés pour dériver des réseaux de drainage de surface à partir de M.N.T<sup>3</sup> (algorithmes  $D8$ ,  $D\infty$ , ...) (Tarboton, 1997; Charleux-Demargne, 2001). Ces modèles peuvent au contraire prendre en compte des incertitudes liées, soit au caractère aléatoire des processus spatiaux, soit à l'incomplétude des données, soit aux données elles-mêmes, e.g. leur localisation et leurs valeurs (Paoli, 2005). Seules les approches stochastiques spatiales formalisées par les approches probabilistes offrent un cadre formel suffisamment développé pour la simulation spatiale numérique (pas de méthode de simulation spatiale en logique floue).

**2.2.3.1.2 Simulation spatiale stochastique** En l'absence totale de données mais connaissant les paramètres et formes des modèles de structure spatiale du champ étudié, on peut réaliser des simulations de ce champ. Ce n'est pas le cas de l'estimation spatiale qui est une méthode d'interpolation de données.

<sup>3</sup>Modèle Numérique de Terrain : on applique des schémas physiques d'écoulements souvent convergents suivant des critères d'altitude entre éléments finis voisins



Les méthodes de simulations spatiales stochastiques s'appuient généralement sur des prototypes de modèles probabilistes dérivés des processus ponctuels spatiaux, de modèles multi-gaussiens géostatistiques (Lantuejoul, 2002), ou des champs aléatoires markoviens (Guyon, 1992). Elles mobilisent également des algorithmes, pour la simulation de distributions standards (gaussiennes, uniformes, ...), pour simuler la structure spatiale (itérations markoviennes, bandes tournantes, décomposition de la matrice de covariance, ...) (Lantuejoul, 2002).

Prenons pour illustration le cas simple en 1D d'un modèle géostatistique stationnaire multi-gaussien qui décrit la variabilité spatiale d'un variable  $Z(s)$  sur une droite ( $s$  désigne un site sur cette droite), avec une structure spatiale de type exponentielle. Autrement dit, ce modèle décrit une variable dont la distribution est gaussienne de loi  $N(\mu, \sigma^2)$ , centrée sur une valeur constante  $\mu$ , et dont la covariance spatiale est une fonction exponentielle décroissante en fonction de la distance  $h$  entre sites sur la droite.

Une simulation issue de ce modèle géostatistique respecte les deux propriétés suivantes :

- En moyenne, tout couple de sites sur la droite suit la fonction de covariance exponentielle, i.e. la différence de leurs valeurs, au carré, est une fonction exponentielle de leur distance ;
- L'histogramme des valeurs simulées dans le champ reproduit la densité de la distribution gaussienne.

Sur la figure 2.7, en partie haute, on peut voir 5 exemples de valeurs simulées tous les 5 m le long d'une droite de 500 m, pour laquelle la fonction de covariance peut s'écrire  $C(h) = \exp\frac{-h}{200}$ ,  $h$  étant une distance sur la droite. La distribution de  $Z$  suit une gaussienne centrée réduite.

L'emploi de modèles spatiaux stochastiques "n'a pas comme ambition de reproduire les mécanismes qui génèrent la variabilité du système ou champ étudié mais plus modestement de mimer une variabilité spatiale aussi réaliste que possible" (Chiles et Delfiner, 1999, p.453).

Cependant, différents algorithmes de simulations peuvent reproduire les mêmes covariances spatiales et les mêmes distributions (histogrammes). La figure 2.8 présente des exemples de simulations sur une droite réalisées à partir de différents algorithmes, différents modèles mathématiques. Ces exemples ont tous la même fonction de covariance spatiale exponentielle. De plus, pour les exemples  $e$  et  $f$ , on reproduit le même histogramme (distribution gaussienne). Sur ces exemples, on peut voir des apparences très différentes bien que les propriétés qui conditionnent ces simulations soient les mêmes. Le réalisme d'une simulation doit donc être regardé au delà de ces deux propriétés.

**2.2.3.1.3 Simulation spatiale conditionnelle** On appelle simulation spatiale conditionnelle, une simulation spatiale qui a des propriétés supplémentaires à celles évoquées ci-dessus (covariance, loi de distribution) qui conditionnent, i.e. contraignent les simulations. Ces contraintes peuvent venir par exemple de règles physiques, déterministes, de la structure du support du champs étudié ... Fréquemment, on parle de simulations conditionnelles pour des simulations qui ont également la propriété de respecter les données, i.e. de "passer" par ces données. Elle sont donc conditionnées par les données, pour tendre vers encore plus de réalisme.

Reprenons le même exemple 1D pour illustrer cette propriété. Sur le bas de la figure 2.7, on peut voir 5 exemples de simulation conditionnelle sur la même base

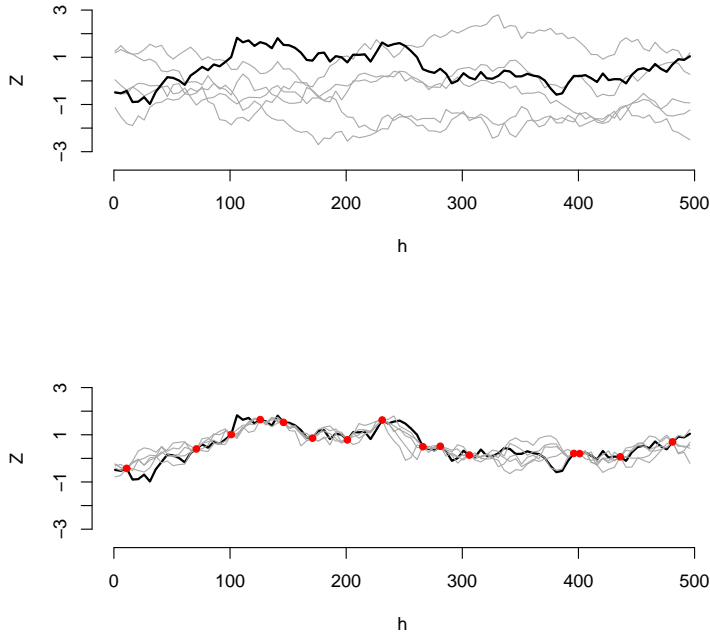


FIG. 2.7 – Haut : Exemples 1D de simulations non conditionnelles (courbes noires et grises) ; Bas : Exemples 1D de simulations conditionnées (courbes noires et grises) par 10 points de données (rouge).

de modélisation (structure de covariance, loi de distribution) que pour la figure du haut, à la différence qu’elles passent par dix points de données (points échantillonnés) dessinés en rouge.

### 2.2.3.2 Simulation conditionnelle vs estimation spatiale

Là encore, restons dans le cadre de la géostatistique. Simulation conditionnelle (conditionnement par des données) et estimation spatiale complètent toutes deux les données manquantes dans l’espace. Elles ont des propriétés communes : elles passent par les données (on parle d’interpolation exacte), et respectent une structure de covariance spatiale. Toutefois, il existe des nuances importantes entre ces deux approches.

L’estimation spatiale se fait par krigeage (simple ou ordinaire) (Cressie, 1993). Pour ce faire, on estime une valeur en un point, ou une petite région, qui : 1– respecte la structure spatiale de la variable, soit une fonction de covariance spatiale (c.f. ci dessus) 2– en imposant un non biais et une variance d’erreur minimum pour cet estimateur. Revenons sur ce dernier critère : variance d’erreur minimum signifie en quelque-sort, que l’on cherche à minimiser l’écart de cette erreur autour de sa moyenne, qui est zéro puisque l’on est non-biaisé. La simulation d’un champ spatial

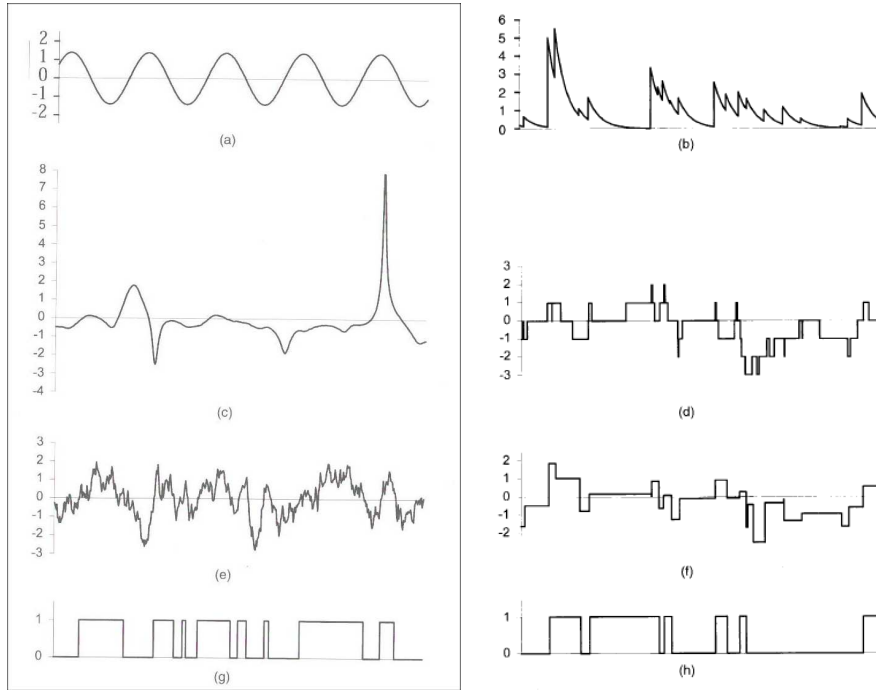


FIG. 2.8 – Différents exemples 1D de simulations présentant la même fonction de covariance spatiale (exponentielle), d'après (Chiles et Delfiner, 1999, p. 454). Les exemples e (simulation d'un vecteur multi-gaussien par la méthode LU) et f (partition poissonnienne en segments gaussiens), ont de plus les mêmes histogrammes.

en géostatistique impose, quant à elle, pour chaque point de l'espace, les conditions suivantes : le champ respecte également une structure spatiale (fonction de covariance spatiale) mais aussi une loi de distribution, par exemple de type multi-gaussienne.

**2.2.3.2.1 Des différences de propriétés locales ou globales** Localement, en un site de l'espace sans donnée, l'estimateur de krigeage propose une valeur qui minimise une variance d'erreur. C'est la meilleure prédiction possible. En revanche, les valeurs issues de plusieurs simulations conditionnelles en ce site n'ont pas cette propriété : elles ont de moins bonnes qualités prédictives. De plus, la qualité prédictive n'est pas équivalente pour ces différentes simulations conditionnelles. La valeur krigée correspond à la moyenne des valeurs de simulations conditionnelles.

A contrario, une propriété bien connue d'un champ de valeurs krigées est d'être plus lisse que le "vrai" champ. Cette propriété de lissage donne des propriétés globales biaisées (histogramme plus resserré), qui peuvent être très fortes si l'on travaille sur des variables dérivées du champ estimé ou sur les extrémités de la distribution de ce champ. En simulation conditionnelle, en revanche, on reproduit également l'histogramme. La variabilité globale du champ est donc plus réaliste, ainsi les variables dérivées de cette variabilité. Un exemple classiquement utilisé pour illustrer cette différence entre estimation et simulation conditionnelle est celui du câble optique sous-marin que l'on souhaite disposer au fond d'un détroit (Chiles et Delfiner, 1999, p.450). On cherche la longueur totale de câble à partir de quelques mesures bathy-

métriques réalisées le long du détroit. Par krigeage, on sous-estime cette longueur alors que par simulations conditionnelles, en moyenne, on reproduit la vraie valeur de longueur totale (voir figure 2.9). Ce biais d'estimation vient du fait que l'on ne travaille pas directement sur la variable modélisée et estimée, la bathymétrie, mais sur une variable dérivée, très sensible aux effets de lissage, puisque la longueur totale intègre les effets de pente du champ bathymétrique.

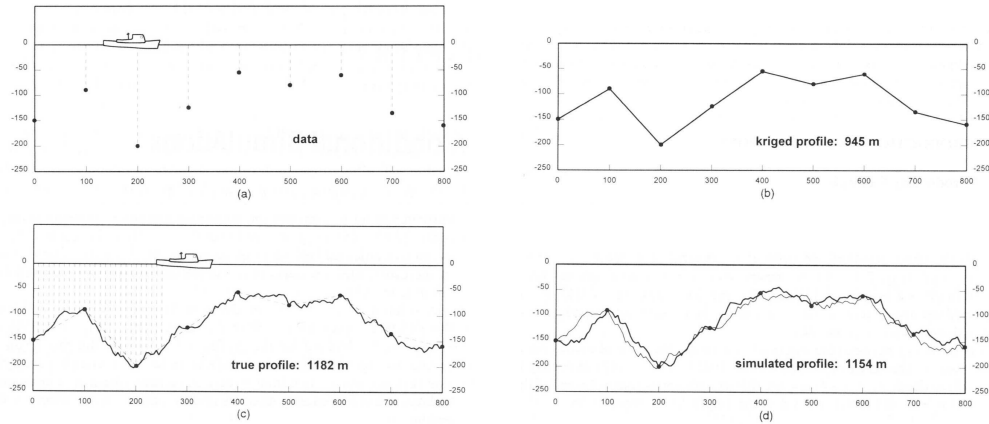


FIG. 2.9 – Illustrations des différences entre estimation et simulation conditionnelle de la longueur de câble sous-marin nécessaire sur un détroit : (a) : sondages ponctuels de la bathymétrie, (c) : vrai champ, (b) : champ estimé par krigeage, (d) : 1 exemple de champ simulé conditionnellement aux points de données, extrait de Chiles et Delfiner (1999, p.450).

**2.2.3.2 Simulation : représentation plus explicite des incertitudes de spatialisation** La cartographie est typiquement le résultat de données ponctuelles et d'estimations combinées. La carte a valeur d'unicité même si un degré d'incertitude peut être associé à chaque valeur estimée (la variance d'erreur de krigeage par exemple), rarement utilisée. Par simulation spatiale conditionnelle, on produit un ensemble de répartitions spatiales globalement équiprobables, qui représente toute la connaissance sur le champ étudié (structure spatiale, distribution, échantillonnage ponctuel), mais surtout toutes les incertitudes sur ce champ. De ce fait, ces représentations spatiales multiples explicitent bien les incertitudes de spatialisation. Elles peuvent être aisément utilisées pour tester la sensibilité d'un modèle qui utilise, par exemple, ce champ spatial comme entrée. Elles permettent donc d'"évaluer quantitativement l'impact des incertitudes de spatialisation sur les résultats de procédures complexes telles la modélisation numérique d'un système dynamique, e.g. un bassin versant cultivé" (Chiles et Delfiner, 1999, p.453).

## 2.3 Démarche méthodologique et objectifs spécifiques

Sur la base des hypothèses introduites et dans une démarche générale de simulation spatiale conditionnelle des réseaux hydrographiques artificiels des paysages cultivés, les objectifs de thèse spécifiques peuvent se décliner en plusieurs points.

Ces différents points correspondent à deux angles de variabilité spatiale des réseaux artificiels :

- la variabilité de propriétés locales du réseau comme par exemple la largeur, la profondeur, le taux d’enherbement... La démarche n’étant que spatiale, on s’intéressera uniquement aux variables stables dans le temps telles les propriétés qui décrivent la gabarit local du réseau. En l’occurrence, comme il a est mentionné dans Bailly (2002), on prendra la largeur du réseau comme variable test des démarches mises en oeuvre et représentative de ces attributs de gabarit.
- la variabilité de l’architecture, i.e. de la géométrie (longueur des rivières, angles, ...) et de la topologie (branchements, nombre de sources...) du réseau, dans l’espace géographique, plus précisément dans l’espace du lattice des limites cadastrales.

Ces deux niveaux de variabilités correspondent aux questions particulières traitées dans les chapitres 3 puis 4 de ce manuscrit, questions déclinées ci-après.

### 2.3.1 Développer des méthodes de simulations et simulations conditionnelles de variables locales sur un arbre orienté

Pour une architecture de réseau fixée, comment modéliser et simuler des variabilités de propriétés locales le long du réseau en tenant compte de la spécificité du support de ces variables (un arbre orienté) ? Comment faire en absence de données ou avec de rares données descriptives ? Comment tenir compte du caractère anthropique de ces variabilités et simuler avec réalisme ce caractère artificiel ?

### 2.3.2 Développer des méthodes de simulations conditionnelles de l’architecture des réseaux

#### 2.3.2.1 Explorer les possibilités de détection de ces réseaux artificiels sur des données LiDAR

Que peuvent apporter des moyens de télédétection pour la détection des réseaux hydrographiques artificiels de bassins cultivés ? Ces réseaux étant des éléments linéaires fins dont la caractéristique principale est de former une discontinuité topographique locale dans le paysage, seules des techniques de télédétection à très haute définition, sensible à ces discontinuités, peuvent être envisagées. En l’occurrence, en pré-ciblant notre analyse sur le lattice des limites cadastrales, quel pouvoir de détection des réseaux, notamment pour les fossés les composent, peut-on espérer en utilisant de la télémétrie Laser aéroportée (LiDAR) ?

#### 2.3.2.2 Simuler une architecture de réseau réaliste à partir d’une connaissance partielle du réseau dans le lattice cadastral orienté par un M.N.T

Si par hypothèse, la télédétection permet une détection partielle dans l’espace de ces réseaux, où si on dispose d’autres informations partielles provenant d’autres

sources de données (carte au 1 :25000 par exemple), comment intégrer cette ou ces informations partielles, incertaines pour simuler un réseau?<sup>4</sup> Comment simuler (interpoler) un réseau arborescent, orienté, à partir de ces segments de réseaux dans le lattice des limites de parcelles cadastrales? Sur quels critères descriptifs juger du réalisme de l'architecture des réseaux simulés?

---

<sup>4</sup>Quelques soient les performances de la télédétection, ce problème se posera toujours : les canalisations enterrées entre fossés ne seront, par exemple, jamais détectées par télédétection.

## 2.4 Présentation du site d'étude

Dans cette section, on présente succinctement les caractéristiques générales du site d'étude retenu, la bassin du Bourdic. Les données sur le site seront précisément décrites au cours des chapitres relatifs à leur utilisation. Dans cette section, nous nous référons principalement aux travaux de Marofi (1999) qui décrit précisément les caractéristiques d'une partie du site d'étude, le bassin de Roujan, caractéristiques que nous extrapolerons à l'ensemble du site étudié.

### 2.4.1 Situation

Le bassin du Bourdic est situé dans le département de l'Hérault en région Languedoc-Roussillon. Il est à cheval entre les communes de Roujan et d'Alignan-du-vent, sur la rive-droite de la partie aval et cultivée du bassin de la rivière Peyne, affluent en rive droite de l'Hérault (confluence à Pézenas). Le bassin du Bourdic inclue le bassin versant de Roujan, bassin d'expérimentations hydrologiques de 91 ha équipé par l'I.N.R.A depuis 1991.

Nous sommes ici au cœur de la plaine viticole languedocienne, vignoble de 280 000 ha (Fig. 2.10).

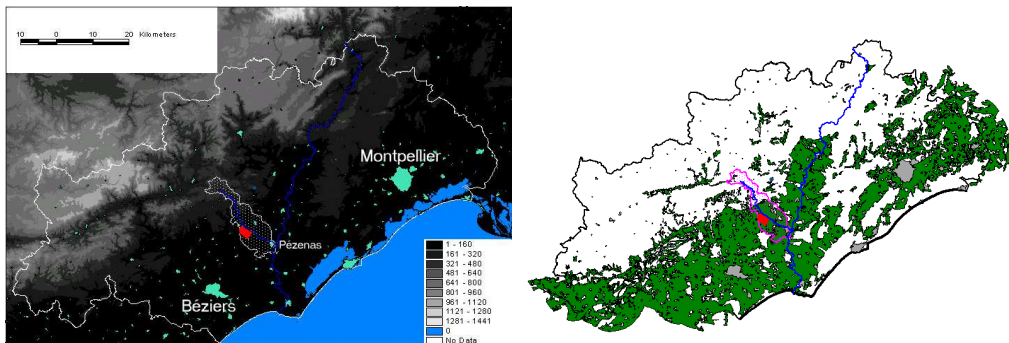


FIG. 2.10 – Situation des bassins de la Peyne (limites grises à gauche, limites roses à droite) et du Bourdic (en rouge) dans le département de l'Hérault. Gauche : relief issu de la Bd-ALTI©IGN et surfaces en vigne (en vert) issues de Corine Land Cover ©IFEN (Droite). Le fleuve Hérault et la Peyne sont en bleu.

Les limites du bassin versant du Bourdic mordent au nord-ouest et sud-est les villages de Roujan et d'Alignan-du-vent et sont donc perturbées par la présence de réseaux d'assainissement urbain sur ces secteurs. On se limitera à la partie cultivée apparente du bassin qui présente une surface d'environ  $6.5 \text{ km}^2$ , dont on peut visualiser une photographie aérienne sur la figure 2.11.

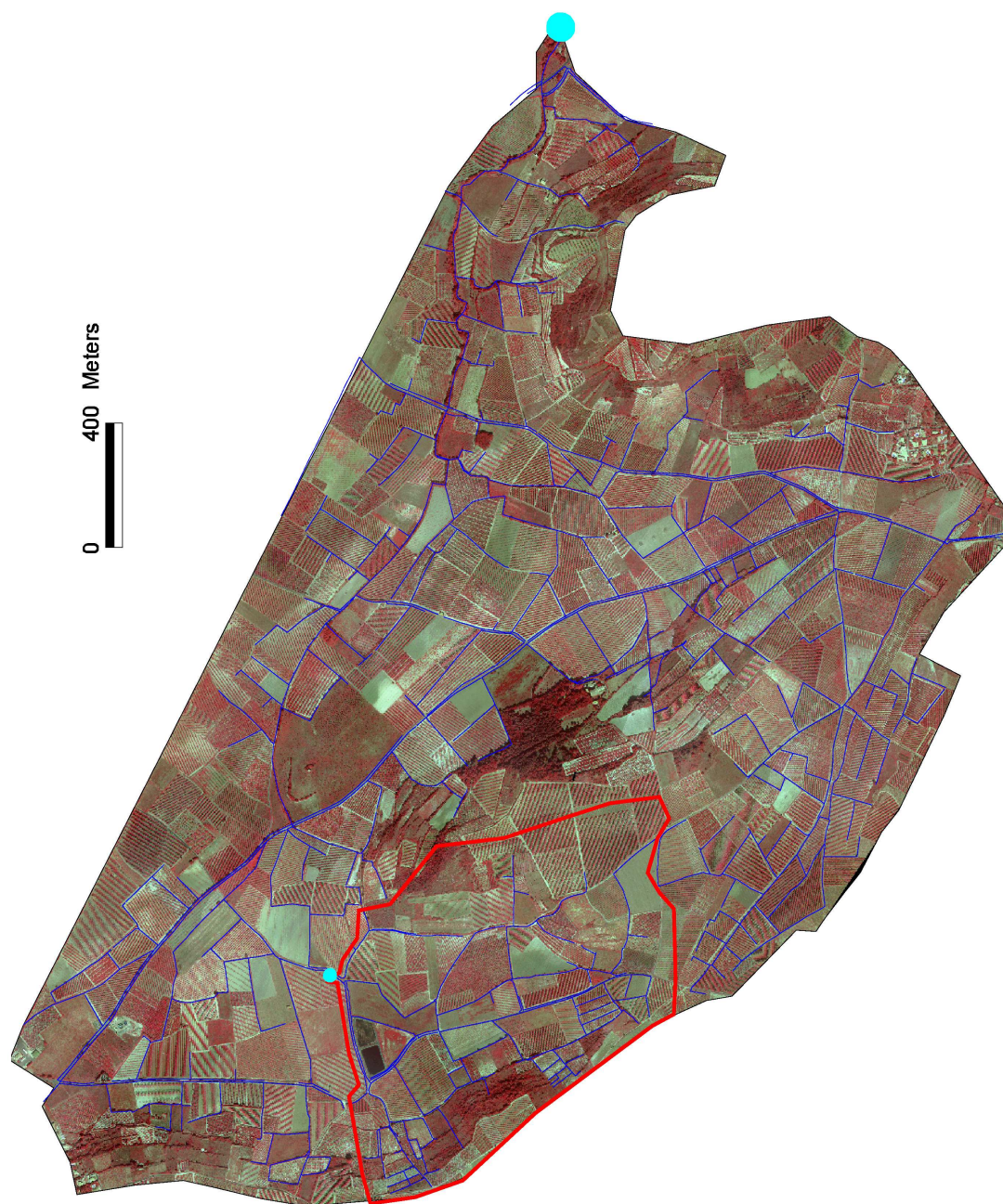


FIG. 2.11 – Image infra-rouge fausses couleurs du bassin du Bourdic incluant le bassin expérimental de Roujan (limite rouge). Le réseau hydrographique, artificiel, est en bleu et les exutoires du Bourdic et du bassin expérimental de Roujan sont marqués par des points en bleu clair.



## 2.4.2 Caractéristiques du milieu

Du point de vue géologique, nous sommes situés sur les bassins sédimentaires du piémont bitérois combinant des dépôts datés de la mer miocène puis de dépôts fluviaux successifs plus tardifs. Cette alternance de dépôts, érosions, en couches géologiques à faible pendage et de perméabilités très différentes, confère au site une organisation en unités géomorphologiques décrites sur la figure 2.12. Ces unités sont classées en unités de plateaux, de terrasses, en unités de coteaux pentus (de plus de 10 % de pente) aménagés de talus (terrassettes), en unités de piémont (glacis) et en unités basses et planes dites de dépression. Ces unités s'organisent sur le site suivant des altitudes (Fig. 2.12). Elles contiennent pour certaines, des nappes perchées (dépression, plateau). Des résurgences de nappes, dites mouillères, sont présentes aux affleurements de couches géologiques de perméabilités différentes.

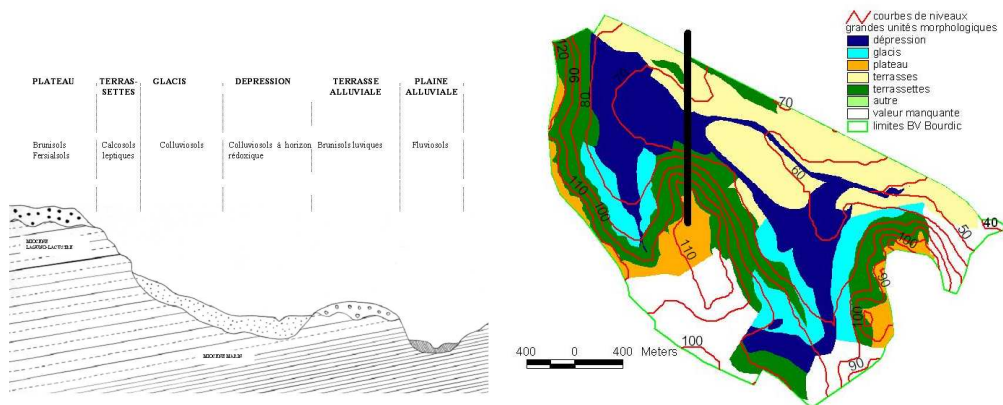


FIG. 2.12 – Gauche : Vue en coupe des unités géomorphologiques du bassin de Roujan, extrait de Lagacherie *et al.* (2001); Droite : unités géomorphologiques cartographiées sur le bassin du Bourdic avec courbes de niveaux (rouge). Les parties blanches sont non renseignées (Source : INRA-LISAH). Le trait noir situe la vue en coupe de gauche.

Du point de vue de l'occupation du sol, la vigne occupe environ 50 % de la surface du bassin et 70 % de la S.A.U<sup>5</sup> du bassin (source : Delenne (2006)). Les viticulteurs sur le site d'étude sont regroupés principalement en deux caves coopératives, celles de Roujan et d'Alignan-du-vent. Compte tenu de la crise du secteur viticole, cette surface en vigne tend à évoluer vers des céréales dans les zones planes et hydromorphes (dépression), vers de nouveaux plantiers de vigne sur les sols les plus favorables (terrasses et plateaux villafranchiens) ou vers de la friche sur les zones les moins favorables aux cultures mécanisées, avec de petites parcelles, i.e. sur les coteaux. La forte croissance démographique de la région tend aussi au développement des zones urbanisées.

<sup>5</sup>Surface Agricole Utile

### 2.4.3 Fonctionnement hydrologique

Le site d'étude est soumis au climat méditerranéen sub-humide à saison sèche prolongée, caractérisé par une grande variabilité inter et intra-annuelle. La pluviométrie totale annuelle moyenne est d'environ 800 mm (sources : Météo France Gabian, station de Roujan INRA-LISAH), répartie principalement entre les mois de septembre à décembre sur quelques événements (des pluies en 24 h de plus de 100 mm ne sont pas extrêmement rares).

Le déficit climatique est important puisque, relativement à ces 800 mm de pluie, on mesure une E.T.P<sup>6</sup> d'environ 1000 mm (source : Météo France Fréjorgues, station de Roujan INRA-LISAH), E.T.P renforcée par les vents marqués sur cette région (Tramontane, vent sec et froid, un tiers de l'année; Marin, vent doux et humide, un cinquième de l'année).

Marofi (1999) distingue trois types d'événements pluviométriques qui marquent trois périodes du cycle hydrologique annuel sur ce site :

- des événements d'automne dits "transitoires (octobre-décembre), importants en hauteur d'eau, qui marquent le début du remplissage des nappes après une longue période sèche,
- des événements d'hiver de faible hauteur, assez nombreux. Ces événements participent au remplissage des nappes et provoquent leur drainage, lorsque le niveau des nappes est supérieur à celui des fossés. On obtient alors des écoulements de base dans les fossés, provenant du tarissement des nappes (notamment des zones basses hydromorphes). A l'exutoire du bassin de Roujan (91 ha), ce débit de base est de l'ordre de  $80 \text{ l.s}^{-1}$  et persiste jusque mai environ.
- des événements d'été "moyens" de juin à septembre, lorsque le déficit hydrique marqué se traduit par une position basse des nappes. Ces événements ne produisent pas d'écoulements à l'exutoire (Roujan) et la lame d'eau précipitée est soit stockée dans le sol soit infiltrée dans les nappes sans atteindre l'exutoire, notamment, via le réseau de fossés (Marofi *et al.*, 2002a).

Une particularité du fonctionnement hydrologique du site vient du système de remplissage/vidange des nappes qui est hétérogène dans l'espace et fortement conditionné par la présence du réseau de fossés (il n'y a pas de drains enterrés récents sur le site). Ce remplissage s'effectue préférentiellement vers les nappes des unités basses de dépression par l'intermédiaire du réseau de fossés qui 1- ré-infiltre les eaux du ruissellement concentré<sup>7</sup> dans ces nappes (Dages, 2006) et/ou 2- ré-infiltre les eaux de drainage des nappes du plateau vers ces nappes (Marofi *et al.*, 2002b).

### 2.4.4 Les différents sites tests utilisés

Sur la bassin du Bourdic, on dispose d'un relevé exhaustif du dessin du réseau hydrographique artificiel dans l'espace géographique, relevé effectué entre 2002 et 2004. Ces relevés ont été numérisés sous S.I.G sur un référentiel cadastral pour aboutir au réseau représenté en bleu sur la figure 2.11. Ce réseau présente un linéaire total de 70 km dont 93% de fossés, 2% de canalisations enterrées, et 5% d'éléments de voiries.

On utilisera dans les chapitres suivants successivement 4 sites tests, sous-parties du bassin du Bourdic, suivant la disponibilité des données utilisées (observations de

---

<sup>6</sup>évapo-transpiration potentielle

<sup>7</sup>produites sur les parcelles agricoles dans un rapport de 1 à 2 suivant le type d'entretien du sol (Chahinian, 2004)

terrain et données de télédétection). Ces sites sont présentés sur la figure 2.13 :

- Le site **A**, correspondant au bassin de Roujan, est celui utilisé dans le chapitre 3. Sur ce site on utilisera toujours le même réseau, mais avec des jeux de données différents sur une propriété locale du réseau. Ces jeux de données, notés **A1** et **A2** sont différents du point de vue du mode, du support (ponctuel ou par segments) et de la date d'échantillonnage de cette propriété, la largeur du réseau.
- Les sites **B** et **C** sont utilisés pour les travaux de télédétection dans les sections 4.2.1 et 4.2.2.
- Le site **D** est utilisé comme site test pour la simulation de l'architecture des réseaux dans les sections 4.3.

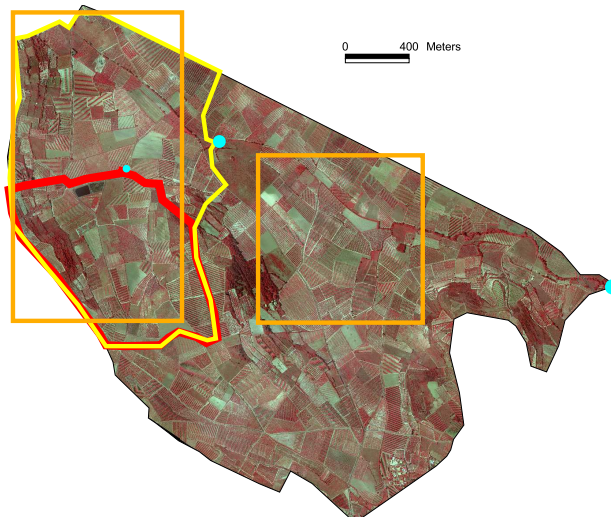


FIG. 2.13 – Limites des différents sites tests utilisés sur le bassin du Bourdic : **A** : Rouge, **B-C** : Orange, **D** : jaune.

# Chapitre 3

## Simulations de variables à support arborescent et orienté

### Sommaire

---

<b>3.1</b>	<b>Introduction . . . . .</b>	<b>53</b>
<b>3.2</b>	<b>Modélisation et simulation géostatistiques de variables spatiales le long d'arbres orientés . . . . .</b>	<b>54</b>
3.2.1	Models and methods . . . . .	54
3.2.1.1	Definitions and notations on drainage network . . . . .	54
3.2.1.1.1	Vertices and edges . . . . .	54
3.2.1.1.2	Directed paths and rivers . . . . .	56
3.2.1.1.3	Edge valuation, lengths and ordering . . . . .	56
3.2.1.1.4	The directed tree as a continuous support of the spatial process . . . . .	56
3.2.1.2	Geostatistical models on a drainage network . . . . .	58
3.2.1.2.1	Spatial distribution, conditional independence and stationarity . . . . .	58
3.2.1.2.2	Weak stationarity . . . . .	60
3.2.1.2.3	Drift models . . . . .	61
3.2.1.3	Experimental variogram . . . . .	61
3.2.1.4	Simulating a random function on a drainage network . . . . .	62
3.2.2	Case study . . . . .	62
3.2.2.1	Data . . . . .	62
3.2.2.2	Description of the data set . . . . .	63
3.2.3	Data modelling and simulation . . . . .	65
3.2.3.1	A test for the absence of spatial structure . . . . .	65
3.2.3.2	Drift modelling . . . . .	66
3.2.3.2.1	Location index for upstream-downstream drift . . . . .	66
3.2.3.3	Residuals variogram modelling and estimation . . . . .	67
3.2.3.3.1	Variogram estimation . . . . .	67
3.2.3.3.2	Variogram fitting . . . . .	67
3.2.3.4	Simulation of $Z(s) = m(s) + Y(s)$ . . . . .	67

3.2.3.5	Computational aspects . . . . .	68
3.2.4	Discussion and conclusions . . . . .	68
<b>3.3</b>	<b>Modélisation et simulation par processus ponctuel de variabilités spatiales par segments homogènes indépendants sur un arbre orienté . . . . .</b>	<b>71</b>
3.3.1	Modèles Poissonniens de changement de segment . . . . .	71
3.3.1.1	Notations . . . . .	71
3.3.1.1.1	Un arbre orienté réalisation d'un processus ponctuel . . . . .	71
3.3.1.1.2	Un arbre orienté support d'un processus ponctuel somme d'un sous-processus fixe et d'un sous-processus aléatoire indépendants . . . . .	71
3.3.1.2	Modèles . . . . .	72
3.3.1.2.1	Modèle poissonnien homogène . . . . .	72
3.3.1.2.2	Modèle poissonnien hétérogène . . . . .	72
3.3.1.3	Estimateurs . . . . .	72
3.3.1.3.1	Estimation de l'intensité $\theta_X$ . . . . .	73
3.3.1.3.2	Estimation de la fonction de Ripley : $K(h)$ . . . . .	73
3.3.1.3.3	Effet de bord : sources et exutoire . . . . .	73
3.3.1.3.4	Estimation aux confluences . . . . .	74
3.3.1.4	Méthodes de simulation . . . . .	74
3.3.1.4.1	Modèle homogène de tessellation . . . . .	74
3.3.1.5	Lien entre processus poissonnien de changement de segment homogène et variogramme . . . . .	75
3.3.2	Cas d'étude . . . . .	77
3.3.2.1	Données . . . . .	77
3.3.2.2	Résultats . . . . .	77
3.3.2.2.1	Modélisation d'une dérive amont-aval de la variable largeur . . . . .	77
3.3.2.2.2	Test de corrélation spatiale du résidu . . . . .	79
3.3.2.2.3	Etude du processus de changement de bief . . . . .	80
	Dérive Amont-Aval de l'intensité . . . . .	80
	Test de $K(r)$ . . . . .	81
3.3.2.2.4	Simulations de sections homogènes à distribution log-normale indépendante . . . . .	81
3.3.2.2.5	Liens entre paramètres du modèle poissonnien et paramètres du variogramme . . . . .	81
3.3.3	Discussion . . . . .	83
<b>3.4</b>	<b>Bilan . . . . .</b>	<b>86</b>
3.4.1	Conditions et limites d'application des modèles . . . . .	86
3.4.2	Validation . . . . .	86
3.4.2.1	Paramètres des modèles . . . . .	86
3.4.2.2	Réalisme des simulations . . . . .	87
3.4.3	Extensions possibles de ces approches de simulation . . . . .	87
3.4.4	Simulations ou simulations conditionnelles? . . . . .	87
3.4.5	Complémentarité et additivité des modèles . . . . .	89

---

## 3.1 Introduction

---

*Ce chapitre est consacré au développement et à l'application de méthodes qui permettent de simuler des variabilités locales le long d'un réseau hydrographique artificiel. Nous proposons deux méthodes présentées chacune dans une section.*

*La première méthode, en première section, est une adaptation de modèles géostatistiques à un support arborescent orienté. Cette première section, en anglais, est extraite d'un article publié dans la revue *Mathematical Geology* et intitulé "Modeling spatial variability along drainage network with geostatistics" (Bailly et al., 2006b). Quelques illustrations et développements pédagogiques complémentaires à cet article (en français ou en anglais) seront toutefois référencés en encadré dans le texte en anglais. Le développement méthodologique proposé permet, sous certaines hypothèses introduites du fait de la particularité du support, d'inférer une fonction spatiale aléatoire sur un arbre orienté. Sur la base des modèles construits, on propose une procédure ad-hoc de simulation de variabilités le long du réseau.*

*La méthode présentée en deuxième section est une application d'un processus ponctuel spatial sur un arbre. On s'intéresse dans ce cas à la modélisation et la simulation de variabilités sur le réseau structurées en segments homogènes, appelés biefs, donc fortement artificielles. Le processus ponctuel marque le changement de biefs sur le réseau.*

*Afin de faciliter leur comparaison, ces deux méthodes sont appliquées sur le même réseau, celui du bassin versant viticole de Roujan, et la même variable, la largeur locale du réseau, prise comme variable emblématique. Elles mobilisent toutefois des données échantillonnées de manières totalement différentes. Dans le premier cas, on utilise le jeu de données **A1** sur la largeur du réseau, échantillonnées ponctuellement sur le réseau. Dans le deuxième cas, on utilise le jeu de données **A2** sur la largeur du réseau, échantillonnée suivant une segmentation a priori du réseau.*

*En dernière section, nous discuterons des différences et complémentarités entre ces deux méthodes. Dans un souci de généralisation de l'application de ces méthodes, nous discuterons de leur limites, de leur validation, de leur extension possibles, ainsi que de leur position par rapport aux derniers travaux méthodologiques sur ce thème. Enfin, puisque ce n'est pas démontré dans les deux premières sections, nous discuterons de leur extension dans le cadre de simulations conditionnées par des données.*

---

## 3.2 Modélisation et simulation géostatistiques de variabilités spatiales le long d'arbres orientés

Extrait de "*Modeling spatial variability along drainage network with geostatistics*" (Bailly *et al.*, 2006b).

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 (Guédon *et al.*, 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 (Lantuejoul, 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, Monestiez *et al.* (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 modeling 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<sup>2</sup>) in the south of France (Louchart *et al.*, 2001).

### 3.2.1 Models and methods

#### 3.2.1.1 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.

**3.2.1.1.1 Vertices and edges** The drainage network is first modeled 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 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. 3.1A).

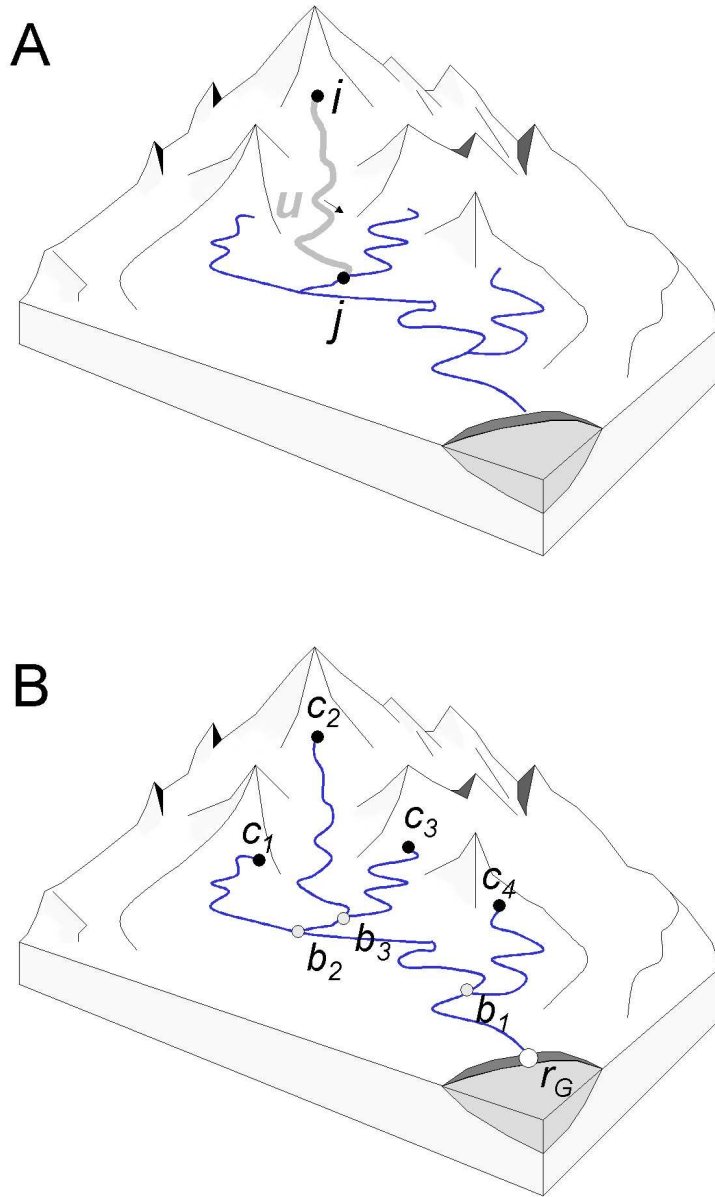


FIG. 3.1 – Example of drainage network. (A) a directed edge  $u = (i, j)$ . (B) the set of sources  $c_1, c_2, c_3, c_4$  (black), the set of junctions  $b_1, b_2, b_3$  (gray) and the outlet  $r_G$  (white)

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. 3.1B).

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. 3.1B).

Vertices with degree  $d_G$  strictly greater than 2 are called *junctions* (Fig. 3.1B).

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

**3.2.1.1.2 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. 3.2-A). 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  $L$  aggregates an edge sequence. We expand on the definition of a directed path in appendix p.169.

Edge  $v$  is a downstream edge of  $u$  if there exists a path containing  $u$  and  $v$  with  $u$  before  $v$ . Let  $\Gamma_u$  be the set of all edges that are downstream edges of  $u$ . Similarly,  $v$  is an upstream edge of  $u$  if  $u$  belongs to  $\Gamma_v$ .  $\Gamma_u^{-1}$  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. 3.2-B) 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  $N_S$ .

**3.2.1.1.3 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 :

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

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 figure 3.1-B. The  $A_o$  set becomes an ordered sequence of rivers :

$$A_o = \left( a_{(1)}, a_{(2)}, \dots, a_{(N_S)} \right), \quad (3.1)$$

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 p.169.

**3.2.1.1.4 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

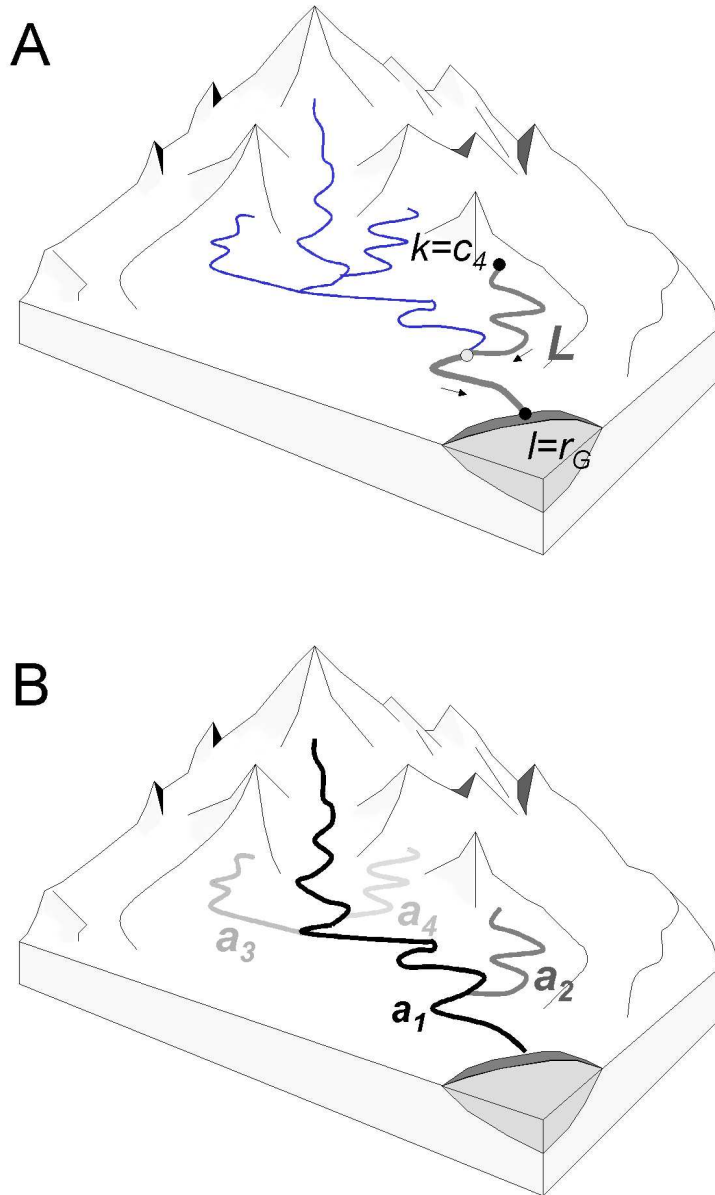


FIG. 3.2 – Example of directed tree. (A) a directed path  $L = (k, l) = (c_4, r_G)$ . (B) a set of rivers  $A_o = \{a_1, a_2, a_3, a_4\}$ .

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. 3.3-A), 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

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

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

If  $u \notin \Gamma_v \cup \Gamma_v^{-1}$ , 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. 3.3-B) :  $x \in u$  with  $u = (i, j) \in U$ ,

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

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.

### 3.2.1.2 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

$\forall z_1, \dots, z_n, \forall s_1, \dots, s_n$ , and  $\forall s'_1, \dots, s'_n$  such as  $s'_1 \in \Gamma_{s_1}, \dots, s'_n \in \Gamma_{s_n}$  and such as  $d(s_1, s'_1) = \dots = d(s_n, s'_n) = h$ , the two probabilities are equal :

$$P\left(Z(s_1) < z_1, \dots, Z(s_n) < z_n\right) = P\left(Z(s'_1) < z_1, \dots, Z(s'_n) < z_n\right), \quad (3.2)$$

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

Equation (3.2) leads to identical random variables  $z(s)$  at any points that are equidistant from the outlet. For example, in Figure 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 (3.2) to this example, it becomes

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

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.

#### 3.2.1.2.1 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 Figure 3.4. Let  $\Gamma_{s_2 s_3}$  be the intersection of  $\Gamma_{s_2}$  and  $\Gamma_{s_3}$ , i.e. the set of downstream points common to  $s_2$  and  $s_3$ .

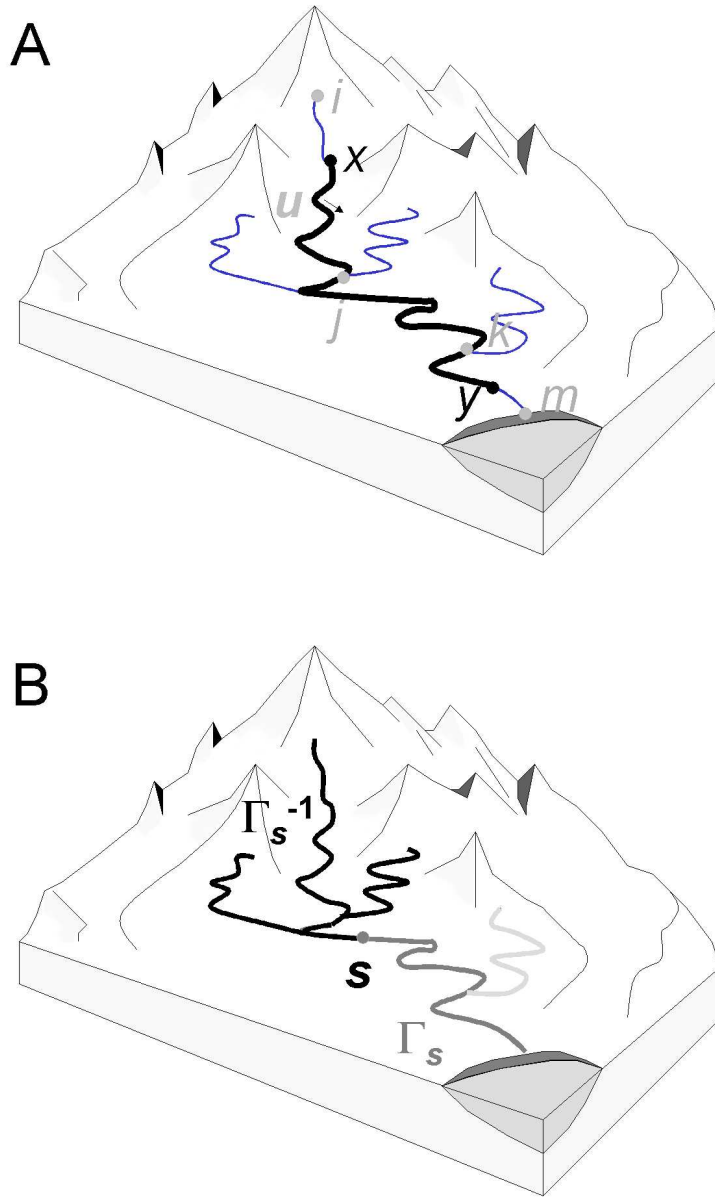


FIG. 3.3 – Example of directed tree. (A) Distance definition on  $G_M$  : distance  $d(xy)$  between two points  $x$  and  $y$  that belong to a same water course. (B) illustration of sets  $\Gamma_s$ , set of downstream points (dark grey),  $\Gamma_s^{-1}$ , set of upstream points (black), and the complement on  $G_M$  to  $\Gamma_s^{-1}$  and  $\Gamma_s$  (light grey), the set of "parallel rivers" for a point  $s$ .

If  $Z(\Gamma_{s_2s_3})$  denotes the random function  $Z$  for all points that belong to  $\Gamma_{s_2s_3}$ ,

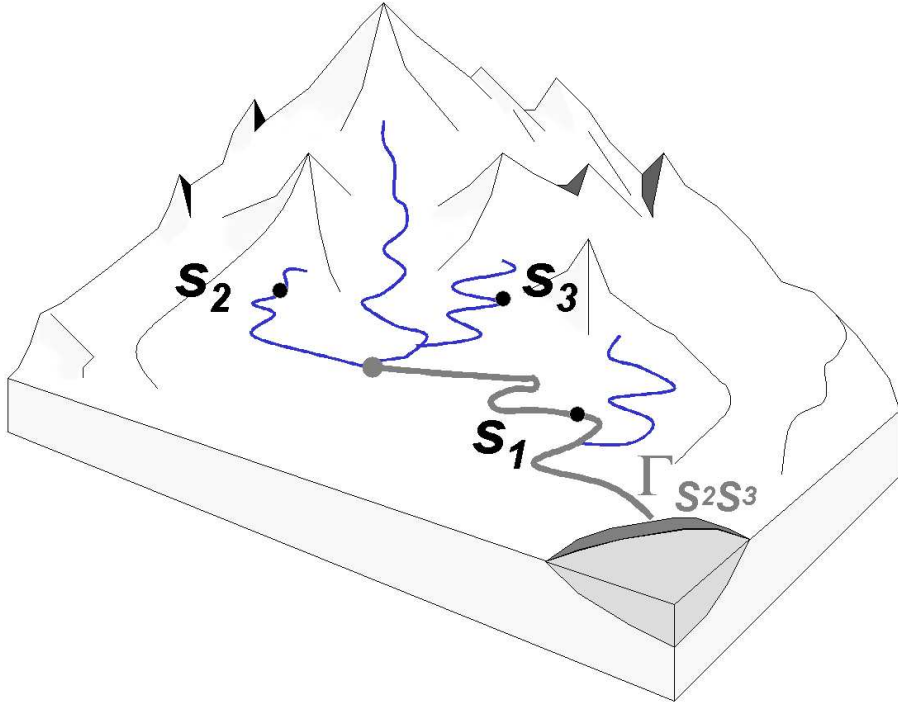


FIG. 3.4 – Stationarity on  $G_M$  : Example of points  $s_2, s_3$  at equal distance from a same downstream point  $s_1$  and  $\Gamma_{s_2s_3}$  in light grey, set of common downstream points of  $s_2, s_3$ .

then the two conditional distributions are equal :

$$\mathcal{L}\left(Z(s_2)|Z(s_3), Z(\Gamma_{s_2s_3})\right) = \mathcal{L}\left(Z(s_2)|Z(\Gamma_{s_2s_3})\right) \quad (3.3)$$

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 (see appendix p. 170) and corresponds to invariance by translation upstream-downstream , i.e. invariance by translation on water courses.

**3.2.1.2.2 Weak stationarity** When the previous hypothesis is limited to only for the first two moments, we obtain :

$$\forall x \in G_M, \mathbf{E}[Z(x)] = m, \quad (3.4)$$

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

$$\text{Cov}(Z(s), Z(s')) = E[Z(s)Z(s')] - m^2 = C(h). \quad (3.5)$$

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

The covariance  $\text{Cov}(Z(s), Z(s'))$  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

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

Similarly, we define the variogram :

$\forall s, s' \in G_M^2$  such as  $s' \in \Gamma_s$  and  $d(s, s') = h$ ,

$$\gamma(h) = \frac{1}{2} E[(Z(s) - Z(s'))^2]$$

**3.2.1.2.3 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 :

$$\forall s_\alpha, \quad Z(s_\alpha) = m(s_\alpha) + Y(s_\alpha) \quad (3.6)$$

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.

### 3.2.1.3 Experimental variogram

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

$$\gamma^*(h) = \frac{1}{2 N_h} \sum_{s_\alpha - s_\beta \simeq h} \left( Z(s_\alpha) - Z(s_\beta) \right)^2 \quad (3.7)$$

where  $N_h$  denotes the count of pairs  $(s_\alpha, s_\beta)$  separated approximatively by the lag  $h$  and such as  $s_\beta$  belongs to  $\Gamma_{s_\alpha}$ .

### 3.2.1.4 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  $s_k, k = 1, \dots, N$  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  $s_k$  (Chiles et 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  $A_o = (a_{(1)}, a_{(2)}, \dots, a_{(N_S)})$  previously defined in equation (3.1).

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 are able to 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  $a_{(i)}$ .

*Voir en annexe p.169 la procédure de simulation détaillée.*

The simulation of all points is completed when points on the last river  $a_{(N_S)}$  have been simulated.

## 3.2.2 Case study

Geostatistical models on directed trees have been applied to an artificial drainage network composed mainly of ditches in a 1 km<sup>2</sup> cultivated catchment, called 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 *et al.*, 2001) as seen on figure 3.6-A.

### 3.2.2.1 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 *et al.*, 2004). The geometry and directed topology of the Roujan network defines a graph  $G$ , which contains 62 sources, 52 junctions and 1 outlet.

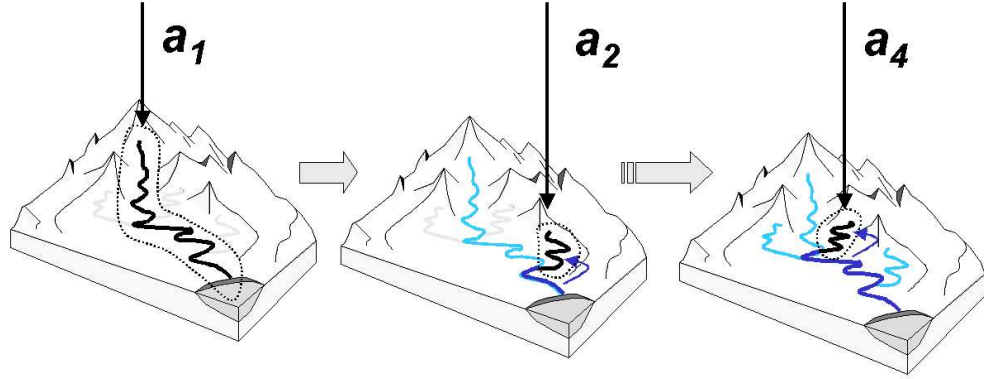


FIG. 3.5 – Simulation of random function on drainage network : procedure is iterated for each river simulation from downstream to upstream : *step 1* - simulation of main river  $a_1$  (black); *step 2* - simulation of river  $a_2$  (black) conditionally to points simulated on  $a_1$  downstream  $a_2$  (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).

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.  $\mathcal{D} = \{s_1, \dots, s_\alpha, \dots, s_N\}$  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.

173 sites, representing 11 % of  $\mathcal{D}$ 's cardinal, were excluded 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.
- pipes : their properties remain constant ; we do not need to model and simulate the spatial distribution of their properties.

Qualitative and quantitative variables have been measured at 1136 sampled sites,  $s_\alpha$ , 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.

### 3.2.2.2 Description of the data set

Figure 3.6-B presents this 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. 3.7). The X coordinate of each vertex  $s_\alpha$ , corresponds to a river number that respects left and right position on



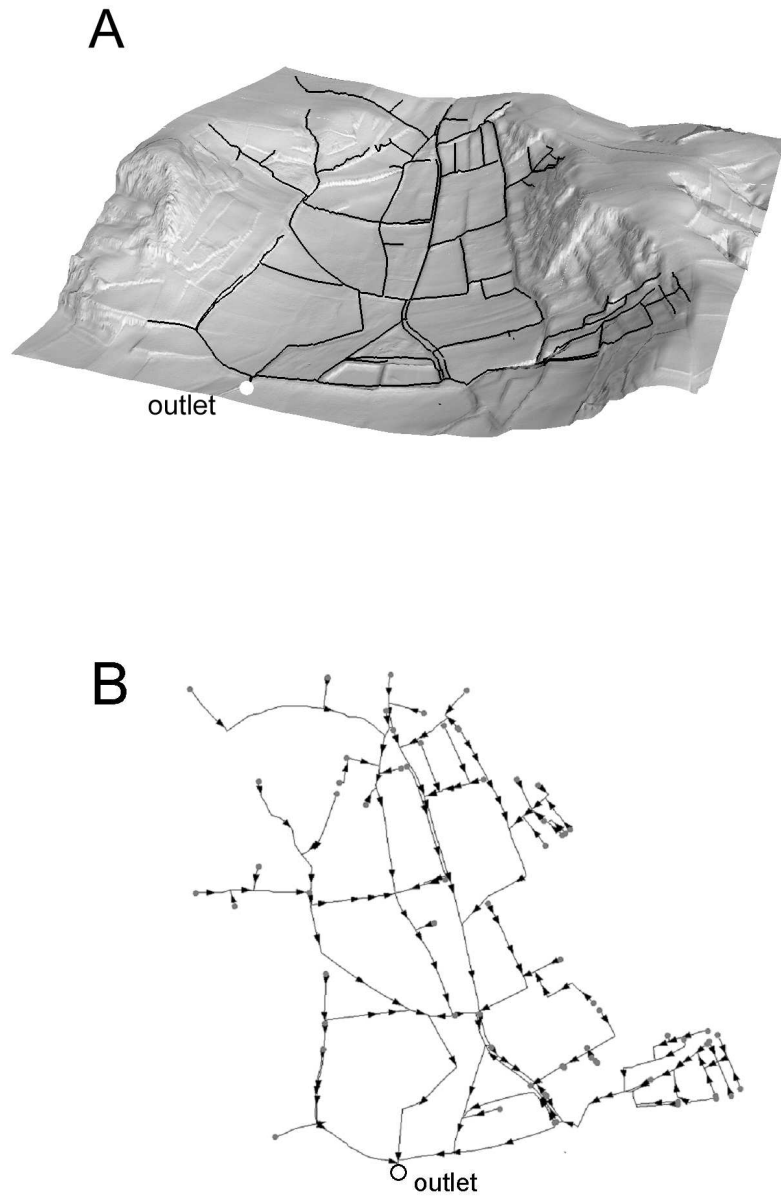


FIG. 3.6 – The artificial drainage network of the Roujan catchment (area of about  $1 \text{ km}^2$ ). (A) depicted in 3 dimensions on a Digital Elevation Model with a vertical scale multiplied by 5. (B) in geographic space (the arrows represent the upstream-downstream direction of each edge).

junctions regarding direction while the Y coordinate corresponds to  $x_1(s_\alpha)$  defined above, the distance to the root.

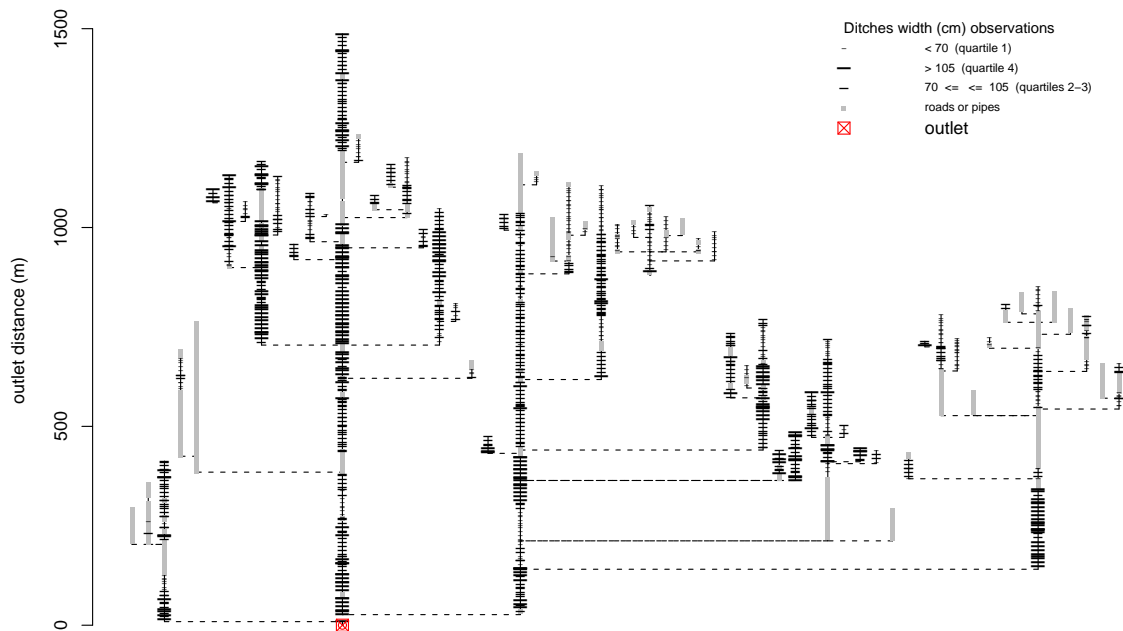


FIG. 3.7 – Width data depicted by quartiles on a topological scheme of the Roujan artificial drainage network : each sample is depicted by a dash with an abscissae given by its position from right to left in the branching structure and with ordinate corresponding to its distance from the root.

Figure 3.7 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.

### 3.2.3 Data modelling and simulation

#### 3.2.3.1 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  $\mathbf{H0}$  which corresponds to complete spatial randomness, i.e. the absence of spatial structure :

$\mathbf{H0}$  :

$$E(Z(s_\alpha)) = m(s_\alpha) = m, \quad \forall s_\alpha \in G_M$$

and independent  $(Z(s_\alpha), Z(s_\beta))$ ,  $\forall (s_\alpha, s_\beta) \in G_M^2$

A Mantel test (Legendre et 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  $E(Z(s_\alpha)) = m(s_\alpha) = f_\theta(x(s_\alpha))$ ,  $\forall s_\alpha \in G_M$ .

### 3.2.3.2 Drift modelling

**3.2.3.2.1 Location index for upstream-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  $s_\alpha$  in a graph. [Eq. (3.8)], as presented above, corresponds to  $x_3$ , the total curvilinear length of edges up to  $s_\alpha$ .

$$x_3(s) = \sum_{v \in \Gamma_u} l(v) + d(i, s) \quad (3.8)$$

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

We first grouped  $s_\alpha$  sites into classes based on  $x_3(s_\alpha)$  upstream-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 3.8 shows on a logarithmic scale for  $x_3(s_\alpha)$   $E(Z(s_\alpha)) = m(s_\alpha) = m$ ,  $\forall s_\alpha \in G_M$ , with empirical drift depicted by points.

Figure 3.8 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  $\theta$  (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  $\hat{y}(s_\alpha) = z(s_\alpha) - m(s_\alpha)$ .

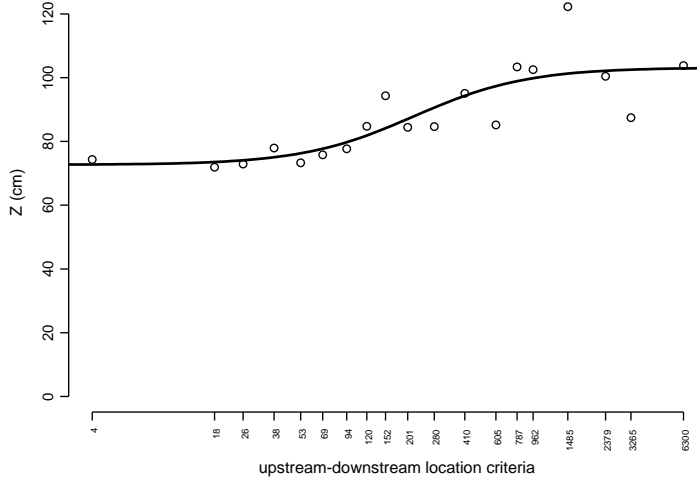


FIG. 3.8 – Experimental (points) and fitted (line) for upstream-downstream drift.

### 3.2.3.3 Residuals variogram modelling and estimation

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

**3.2.3.3.1 Variogram estimation** The variogram was estimated using the classical variogram estimator given in equation (3.7). The analysis was limited to 250 meters pairs distance on  $G_M$ . Figure 3.9 with points corresponding to experimental variogram computed on 10-meter lags<sup>1</sup> 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.

**3.2.3.3.2 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) :

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

with  $\theta = (a, b, c)$ , corresponding classically to sill, range and nugget effect.

The fitted parameters are  $\hat{\theta} = 705.8, 81.9, 87.9$ .

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

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

<sup>1</sup>numbers along the x axis correspond to the number of pairs for each lag

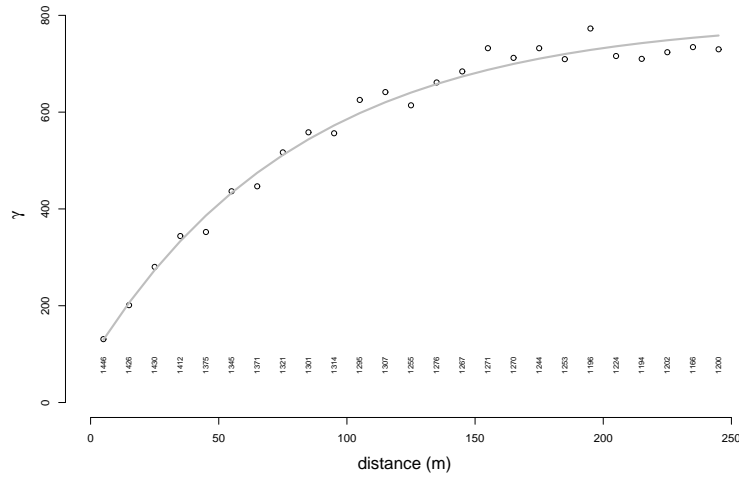


FIG. 3.9 – Experimental (points) and fitted (line) variograms on drift residuals.

*Voir en annexe p. 171 une illustration de la procédure de simulation conditionnelle aval-amont sur une rivière.*

Figure 3.10 (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 Figure 3.10.

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

### 3.2.3.5 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 as well as 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 located 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.

## 3.2.4 Discussion and conclusions

This paper proposes 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

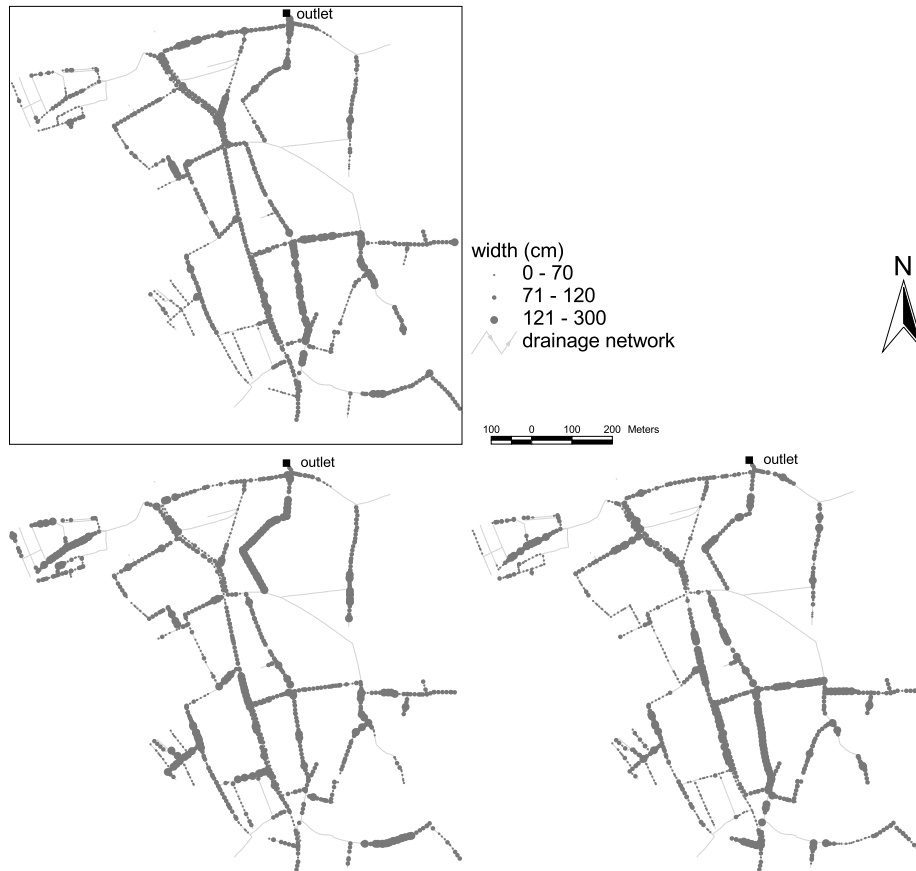


FIG. 3.10 – Width data on top-left depicted by dark grey points with legend on top-right. The 2 sub-figures down are examples of simulated values depicted with the same legend as for the actual ones.

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-downstream drift but high 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 et Maddock (1953). High 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 in order to si-

simulate variables like 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, in order to reproduce reality more closely, the simulations would have to be able to generate homogeneous sections.

It is 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 modeling.

### 3.3 Modélisation et simulation par processus ponctuel de variabilités spatiales par segments homogènes indépendants sur un arbre orienté

Pour modéliser et simuler des variabilités par segments homogènes le long d'un arbre orienté, la méthodologie proposée s'appuie sur un processus ponctuel associé à une loi de distribution d'une variable ayant pour support les segments bornés par ce processus ponctuel. Cette approche de modélisation s'apparente à un modèle de tessellation<sup>2</sup> en 1D (Lantuejoul (2002), p.133), appliqué à un arbre orienté. Dans notre cas, le processus ponctuel ne marque pas le "germe" des segments (usuellement de type Voronoï) mais il marque les extrémités des segments.

Cette approche de modélisation de variabilités spatiales dans un arbre orienté selon des segments homogènes est une alternative de modélisation et de simulation spatiale à la modélisation géostatistique présentée auparavant. En effet, l'homogénéité par segment d'une variable sur un réseau marque bien souvent le caractère fortement artificiel de ce réseau. On verra dans ce chapitre et dans le bilan de fin de chapitre que les approches processus ponctuel et géostatistique sont liées, et peuvent être complémentaires.

#### 3.3.1 Modèles Poissonniens de changement de segment

##### 3.3.1.1 Notations

**3.3.1.1.1 Un arbre orienté réalisation d'un processus ponctuel** Soit l'arbre orienté et valué  $G = [D, U_G]$ , avec  $D$  ensemble des noeuds et  $U_G$  ensembles des arêtes. Comme pour les modèles géostatistiques, chaque arête  $u = (i, j)$  correspond à un segment de réseau hydrographique naturellement orienté de l'amont vers l'aval ayant un noeud amont  $i$  et un noeud aval  $j$ . On note également  $S$  l'ensemble des sources,  $B$  l'ensemble des confluences et  $r_G$  l'exutoire qui, assemblés, forment l'ensemble des noeuds du graphe  $G$  :

$$D = S \cup B \cup r_G$$

On parlera ensuite pour les noeuds de  $D$ , fixés par la structure topologique de  $G$ , d'une réalisation fixe d'un processus  $\mathbf{D}$  pouvant être considéré comme aléatoire quand  $G$  lui-même, donc la topologie du graphe, est aléatoire. Dans notre cas, on considère  $\mathbf{D}$  comme une réalisation fixe de ce processus général aléatoire.

**3.3.1.1.2 Un arbre orienté support d'un processus ponctuel somme d'un sous-processus fixe et d'un sous-processus aléatoire indépendants** On désigne par  $G_P = [X, U]$  l'arbre orienté (équivalent de l'arbre  $G_M$  dans la section précédente) support d'un processus ponctuel noté  $\mathbf{X}$ .  $G_P$  a la même structure topologique que  $G$ .  $G_P$  est également l'espace support d'une variable étudiée  $Z$  définie en tout site  $s$  de  $G_P$ .  $Z$  est répartie en segments homogènes formant l'ensemble  $U$  des arêtes de  $G_P$ , appelées *biefs*.

Chaque bief  $u \in U$  a pour extrémité initiale un noeud  $i \in X$ , étant soit une source, soit une confluence, soit ce que l'on appelle un *événement*, i.e. un noeud de

---

<sup>2</sup>division de l'espace en cellules ou unités homogènes



degré  $d = 2$  étant ni une source, ni une confluence et un noeud  $j \in X$ , étant soit une confluence, soit la racine, soit un événement.

$X$ , l'ensemble des noeuds de  $G_P$  désigne la réalisation du processus ponctuel  $\mathbf{X}$ . On le suppose dans notre cas comme la somme de 2 sous-processus ponctuels indépendants :  $\mathbf{D}$  et  $\mathbf{A}$ .

$\mathbf{D}$ , vu précédemment est une réalisation d'un sous-processus aléatoire fixé par la topologie du graphe  $G_P$ , avec  $D = S \cup B \cup r_G$ , ensemble réuni des sources, confluences et exutoire. L'ensemble des noeuds "événements", marquant un changement de bief pour la variable étudiée  $Z$  entre deux confluences, une source et une confluence, . . . , est une réalisation du sous-processus aléatoire  $\mathbf{A}$ .

$\mathbf{A}$ , aléatoire, est un processus que l'on considère comme indépendant de  $\mathbf{D}$ .

On a donc par construction  $X = S \cup B \cup r_G \cup A$ . On définit également sur  $G_P$  des chemins orientés, des rivières, une valuation des arêtes par leurs longueurs.

La variable étudiée  $Z$  est donc constante sur chaque bief et chaque point du processus ponctuel  $\mathbf{X}$  marque une extrémité de bief, donc une variation très probablement non continue de  $Z$ . Par convention,  $Z$  est constante sur chaque bief à l'exclusion du noeud aval. Si  $z_1$  est une valeur constante de  $Z$  sur  $u = (i, j)$  alors

$$\forall u = [i, j] \in U, \forall s_\alpha, s_\beta \in [i, j]^2, Z(s_\alpha) = Z(s_\beta) = z_1 \quad (3.9)$$

On note  $N_X$  la fonction de comptage du processus  $\mathbf{X}$ ,  $N_D$  celle du processus  $\mathbf{D}$  et  $N_A$ , celle du processus  $\mathbf{A}$ .

### 3.3.1.2 Modèles

**3.3.1.2.1 Modèle poissonnien homogène** Si  $\mathbf{X}$  est un processus poissonnien d'intensité  $\theta_X$ , on note, pour tout partie  $g$  de  $G_P$  :

$$P\{N_X(g) = n\} = \exp^{(\theta_X)} \frac{(\theta_X)^n}{n!}$$

L'intensité représente un nombre moyen de points par unité de longueur  $du$  sur l'arbre. Si un processus poissonnien est la somme de deux autres indépendants alors l'intensité résultante est la somme des deux intensités  $\theta_X = \theta_A + \theta_D$ .

**3.3.1.2.2 Modèle poissonnien hétérogène** Dans ce cas,  $\theta_X$  n'est plus constante et si  $s$  désigne un site quelconque de l'arbre support du processus ponctuel  $G_P$ , on note alors cette intensité variable comme une fonction d'un indice de localisation amont-aval de  $s$  sur l'arbre.

$$\theta_X(s) = h(x_i(s))$$

avec  $x_i(s)$  parmi :

- $x_1(s) = d(s, r_G)$ , distance à la racine,
- $x_2(s)$ , distance maximale à une source,
- $x_3(s)$ , le linéaire amont cumulé (cf. partie 3.2.1).

### 3.3.1.3 Estimateurs

La description d'une structure spatiale d'un semis de points se base généralement sur les propriétés des deux premiers moments du processus : l'intensité pour le

premier moment (ici, probabilité de présence d'un point sur un segment élémentaire  $du$ ) ; la propriété de second moment est caractérisée par une fonction de densité de paires de points (voir Goreaud (2000), p.37), notamment au moyen de la fonction de Ripley que nous avons choisi d'utiliser du fait qu'elle soit basée sur les distances, ici distances curvilinéaires orientées dans l'arbre (cf sous-chapitre sur les modèles géostatistiques).

$$\forall x \in X, K(h) = \frac{1}{\theta_X} E[\text{nombre de voisins à distance } \leq h \text{ de } x] \quad (3.10)$$

Tout comme en géostatistique, l'inférence des deux premiers moments du processus se fait suivant des hypothèses d'ergodicité et de stationnarité du processus (repliation par translation amont-aval dans l'arbre du processus permettant de remplacer sa réalisation unique Cressie (1993), p.57).

**3.3.1.3.1 Estimation de l'intensité  $\theta_X$**  L'estimation de la fonction d'intensité locale  $\theta_X(s)$  pour un site quelconque  $s$  de l'arbre  $G_P$  est liée au processus de comptage local  $N_X$ . Sur une droite, on l'estime suivant un noyau gaussien, d'écart-type  $T$  :

$$\hat{\theta}_X(s, T) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}T} \exp\left(-\frac{x^2}{2T}\right) N_X(s-x) dx$$

On utilisera par de même par la suite une fonction à noyau gaussien pour les estimations d'intensités localisées  $\theta_X(s)$  ou  $\theta_A(s)$  sur l'arbre (Cressie (1993), p. 600).

**3.3.1.3.2 Estimation de la fonction de Ripley :  $K(h)$**  Soit  $\hat{\theta}_X$  l'intensité du processus estimée. Un estimateur de la fonction de Ripley devient pour un arbre orienté :

$$\hat{K}(h) = \frac{1}{\hat{\theta}_X} * \sum_{s_\alpha \in X} \sum_{s_\beta \neq s_\alpha \in X} \frac{\mathbf{1}_{d(s_\alpha, s_\beta) \leq h}}{N}$$

avec  $N$ , cardinal de  $X$ ,

et  $d(s_\alpha, s_\beta)$ , distance orientée curvi-linéaire dans l'arbre entre  $s_\alpha$  et  $s_\beta$ .

Cet estimateur est connu pour être biaisé par les effets de bord en limite du domaine d'étude. Ceci reste vrai dans le cas d'un arbre (biais à proximité des sources et de la racine). Un autre problème provient de l'estimation aux confluences. Nous avons choisi de corriger ces effets par les méthodes présentées ci-après. Elles sont présentées pour les estimateurs du processus  $\mathbf{X}$  mais elles sont équivalentes pour  $\mathbf{A}$  ( $\mathbf{D}$  fixé).

**3.3.1.3.3 Effet de bord : sources et exutoire** Les biais d'estimation des intensités  $\theta$ , ou des fonctions de Ripley  $\hat{K}(h)$  liés aux effets de bord proviennent de la non continuité dans l'espace du processus de comptage  $N_X$  (voir Goreaud (2000), p. 50). Dans ce cas, nous proposons de corriger cet effet en prolongeant artificiellement l'espace support par symétrisation depuis la source du chemin amont-aval (Fig. 3.11) ou depuis la racine du chemin aval-amont. Pour ce dernier cas, si une confluence est à proximité de la racine, alors le comptage doit être étendu comme expliqué dans le paragraphe ci-après.

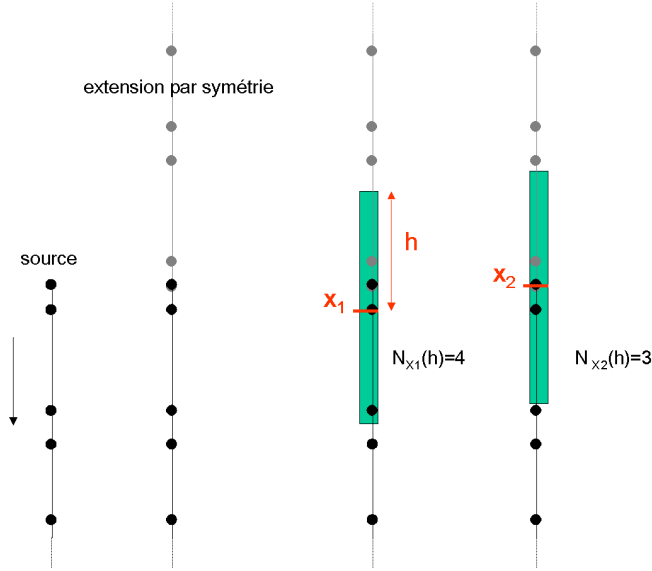


FIG. 3.11 – Correction des effets de bords aux sources et à la racine sur le processus de comptage du semis de points : la boîte verte symbolise l’intervalle de la fenêtre glissante pour le comptage (ou noyau).

**3.3.1.3.4 Estimation aux confluences** De même que précédemment, un processus de comptage particulier doit être introduit pour les estimateurs locaux en aval d’une confluence  $c_i$  du fait de la dimension de l’espace à considérer. Nous proposons ici un processus de comptage qui moyenne simplement les processus de comptage en  $1D$  sur chaque chemin amont-aval passant par la confluence  $c_i$ . Par exemple, pour compter les voisins d’un point situé juste en aval d’une confluence dans une fenêtre de distance  $h$ , on obtient le processus de comptage expliqué en figure 3.12. Ce processus de comptage peut créer des discontinuités (estimations d’intensités, de fonction de Ripley) aux confluences.

#### 3.3.1.4 Méthodes de simulation

La simulation recherchée sous-segmente un arbre déjà segmenté par sa topologie de départ (sources, confluences, exutoire) d’une manière telle que le résultat soit la réalisation d’un processus poissonnien à intensité constante (ou variable). L’attribution de valeurs simulées de  $Z(u)$  sur ces segments dans un deuxième temps ne pose pas de difficulté particulière surtout si elles sont distribuées de manière indépendante.

**3.3.1.4.1 Modèle homogène de tessellation** Soit  $Z$  la variable étudiée définie en tout point  $s$  de l’arbre.  $Z$  est distribuée sur l’arbre en section ou segments homogènes  $u$ . Chaque valeur  $z$  de  $Z$  a donc pour support une arête  $u = [i, j[$  de  $G_P$ .  $\mathbf{X}$  est un processus poissonnien homogène, somme de deux sous-processus poissonniens indépendants  $\mathbf{D}$  et  $\mathbf{A}$  eux aussi homogènes.  $Z$  suit une loi de probabilité  $\mathcal{F}(\mu, \sigma)$  à support  $u$  sans corrélation spatiale. On s’intéresse ici à la procédure de simulation

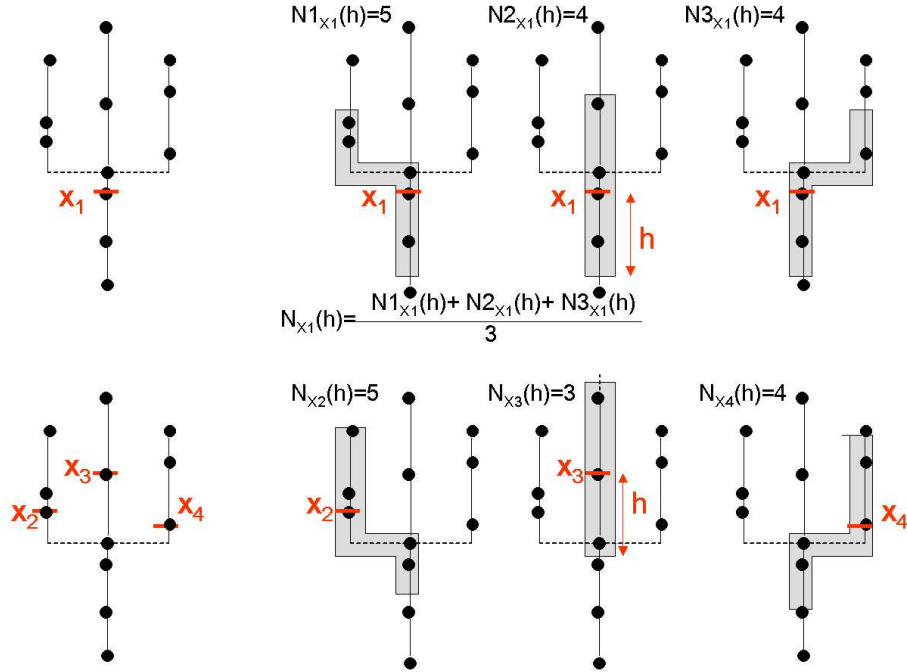


FIG. 3.12 – Processus de comptage du semis de points aux confluences ; Haut : comptage des voisins à la distance  $h$  d'un point situé juste en aval d'une confluence ; Bas : comptage des voisins à la distance  $h$  d'un point situé juste en amont d'une confluence. La boîte grise symbolise l'intervalle de la fenêtre glissante (ou noyau) pour le comptage.

de  $\mathbf{A}$  ou  $\theta_A$ . On considère le sous-processus  $\mathbf{D}$  déjà réalisé (topologie de l'arbre  $G$ ) et indépendant.

Le processus de simulation de  $\mathbf{X}$  ( $\mathbf{A}$ ) puis de  $Z(u)$  est :

1. Démarrer avec  $X = D$
2. Estimer  $\theta_A$  par  $\frac{\text{card}\{A\}}{L}$  avec  $L$  linéaire total de l'arbre
3. Concaténer les unes derrière les autres chaque rivière  $r_j$  de  $A_o$  dans un ordre quelconque pour former un seul support 1D noter  $L_P$  de longueur  $L$
4. Sur  $L_P$  générer  $n(L_P) \sim \text{Poisson}(\theta_A)$
5. Si  $n(L_P) > 0$ , alors pour  $i = 1, \dots, n$  générer  $x_i^{L_P} \sim \text{Unif}(L_P)$ . On note  $A^{L_P}$  l'ensemble  $\{s_1^{(L_P)}, \dots, s_n^{(L_P)}\}$
6. Réunir pour former  $X = D \cup A^{L_P}$
7. Réattribuer les noeuds  $X$  par rivière et former les segments  $u$  par rivière
8. Sur le graphe  $G_P = [X, U]$  résultant, générer alors pour chaque  $u$ ,  $z_k \sim \mathcal{F}(\mu, \sigma)$  indépendamment.

### 3.3.1.5 Lien entre processus poissonnien de changement de segment homogène et variogramme

Dans cette partie, nous allons démontrer que le variogramme pour toute paires de points quelconques pris sur un arbre orienté  $G_P$ , support d'une variabilité spatiale par segments homogènes (biefs) de distribution connue  $\mathcal{F}(\mu, \sigma^2)$  est de forme

exponentielle avec des paramètres connus.

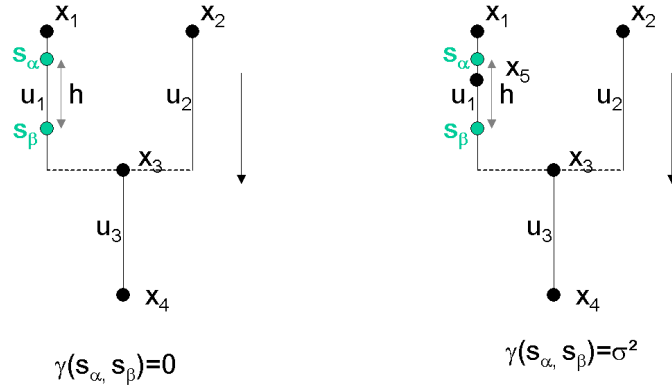


FIG. 3.13 – Gauche :  $s_\alpha$  et  $s_\beta$  appartiennent au même segment, Droite :  $s_\alpha$  et  $s_\beta$  sont sur deux segments différents

Pour tout couple de points  $(s_\alpha, s_\beta)$  distant de  $h$  sur l'arbre  $G_P$  :

$$\gamma(s_\alpha, s_\beta) = \gamma(h) = \frac{1}{2}E[(Z(s_\alpha) - Z(s_\beta))^2]$$

On suppose le changement de bief décrit par un processus ponctuel poissonnien d'intensité  $\theta_X$ .

Si il y a au moins un changement de segment homogène entre  $s_\alpha$  et  $s_\beta$ , i.e.  $s_\alpha$  et  $s_\beta$  sont sur deux segments différents (Fig. 3.13-droite), et comme les segments portent des valeurs de même loi mais indépendantes entre elles (i.i.d), on obtient alors  $\gamma(s_\alpha, s_\beta) = \sigma^2$ . Si il n'y a pas de changement de segment homogène donc si  $s_\alpha$  et  $s_\beta$  appartiennent au même segment homogène (Fig. 3.13-gauche), alors  $\gamma(s_\alpha, s_\beta) = 0$ .

Autrement dit, on peut écrire :

$$\begin{aligned} \gamma(h) &= \frac{1}{2}E[(Z(s_\alpha) - Z(s_\beta))^2] \iff \\ \gamma(h) &= \frac{1}{2}P([s_\alpha, s_\beta] \in \text{même bief}) * 0 + \frac{1}{2}P([s_\alpha, s_\beta] \in 2 \text{ biefs } \neq) * 2 * \sigma^2 \iff \\ \gamma(h) &= P([s_\alpha, s_\beta] \in 2 \text{ biefs } \neq) * \sigma^2 \iff \\ \gamma(h) &= (1 - P([s_\alpha, s_\beta] \in \text{même bief})) * \sigma^2 \iff \\ \text{or } P([s_\alpha, s_\beta] \in \text{même bief}) &= \exp^{-\theta h} \text{ (loi de Poisson)} \end{aligned}$$

donc :

$$\gamma(h) = \sigma^2 * (1 - \exp^{-\theta h})$$

On obtient un variogramme exponentiel dont les paramètres, palier et portée, sont respectivement la variance  $\sigma^2$  et  $\theta_X$ , intensité du processus ponctuel.

### 3.3.2 Cas d'étude

#### 3.3.2.1 Données

Le réseau étudié est le même que celui étudié dans la section précédente 3.2. Il s'agit du réseau de fossés du bassin versant expérimental de Roujan (bassin de 1  $km^2$ , réseau de 11 km). Seules les données sur ce réseau diffèrent. On utilise dans ce cas le jeu de données **A2**. Un relevé terrain des propriétés de ce réseau effectué par l'INRA en 2000 (Diot, 2000), a été réalisé suivant un découpage du réseau en 265 biefs homogènes. La délimitation des biefs est conditionnée par le changement d'occupation du sol d'une des deux parcelles juxtaposant le réseau, les confluences, ou un changement net pour au moins une des propriétés relevées (typologie AQUAE, (Lagacherie *et al.*, 2004)). Ce découpage relève de la variabilité combinée de plusieurs variables. Des données ont été échantillonnées sur 45 à 95 % des biefs : largeur (largeur moyenne entre la largeur au fond du fossé et la largeur au toit du fossé), profondeur, taux d'enherbement... Nous nous intéresserons ici à la variable largeur qui a été échantillonnée sur 174 biefs. Pour chaque bief, on dispose donc d'une valeur observée considérée comme constante sur tout le bief. L'arbre  $G_P$  étudié a les propriétés suivantes : 62 sources, 52 confluences, 1 racine, 265 noeuds et donc 264 arêtes. La figure 3.14 montre l'ensemble des noeuds  $X$  délimitant les biefs et formant le semis de points étudié (en rouge  $A$  et en bleu  $D$ ).

La largeur par bief observée dont l'histogramme est présentée à gauche de la figure 3.16 montre une distribution log-normale.

#### 3.3.2.2 Résultats

##### 3.3.2.2.1 Modélisation d'une dérive amont-aval de la variable largeur

Comme en section 3.2, afin de modéliser une dérive amont-aval a priori de la largeur, nous construisons un critère de position amont-aval des biefs dans le réseau correspondant au cumul du linéaire amont  $x_3(u)$  de chaque bief  $u$  (même critère et même justifications que précédemment). Ce critère de position par bief est calculé arbitrairement pour le noeud amont du bief noté  $s_\alpha$ . On effectue donc une transformation de variable pour tous les biefs  $u$  :

$$Z(u) = m(u) + Y$$

, avec  $m(u) = f(x_3(s_\alpha))$ , dérive amont aval et  $Y$ , résidu par bief. Comme dans la section précédente, si l'on représente l'évolution de la largeur des biefs suivant ce critère de position amont-aval, on obtient sur une échelle logarithmique en  $x$ , la figure 3.15. Cette dérive, de même forme que dans la section précédente, est modélisée par

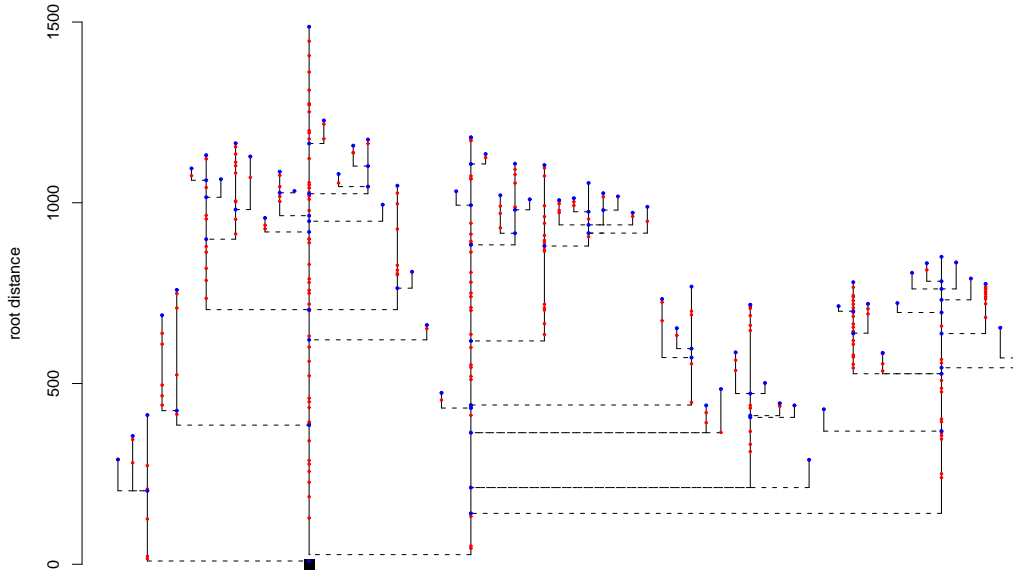


FIG. 3.14 – Le réseau de fossés du bassin de Roujan : les limites de biefs sont marquées par les points (rouge= $A$ , bleu= $D$ ).

la même fonction logistique. Pour tout point  $s_\alpha$ , noeud amont d'un bief dans le réseau, le modèle s'écrit :

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

avec  $p(s_\alpha) = \log(x_3(s_\alpha))$  et  $x_3(s_\alpha)$ , linéaire cumulé en amont de  $s_\alpha$ . Les paramètres ajustés sont :

$$\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \hat{\theta}_4) = (34, 3.3, 4.6, 91)$$

Ces paramètres diffèrent légèrement de ceux de la modélisation de dérive de la partie géostatistique (ce ne sont pas les mêmes données, pas le même support).

On calcule ensuite les résidus par bief de ce modèle de dérive.

Sur la droite de la figure 3.16, on peut visualiser l'histogramme des résidus. On observe toujours une distribution log-normale mais plus stabilisée que pour la largeur  $Z(u)$ .

On ajuste ensuite cette distribution par une loi log-normale de paramètres  $\mu = -5, \sigma^2 = 2638$  représentée en rouge sur la droite de la figure 3.16.

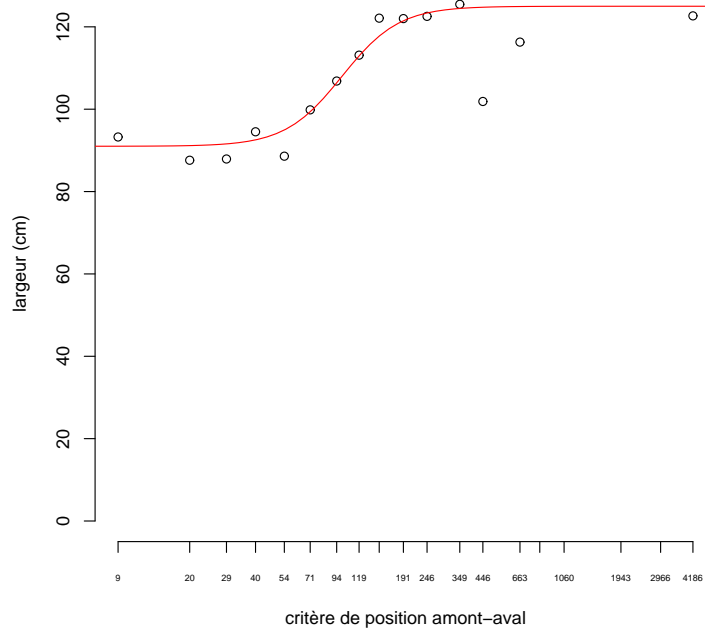


FIG. 3.15 – Evolution de la largeur des biefs suivant un critère de position amont-aval et modèle de dérive ajusté (rouge)

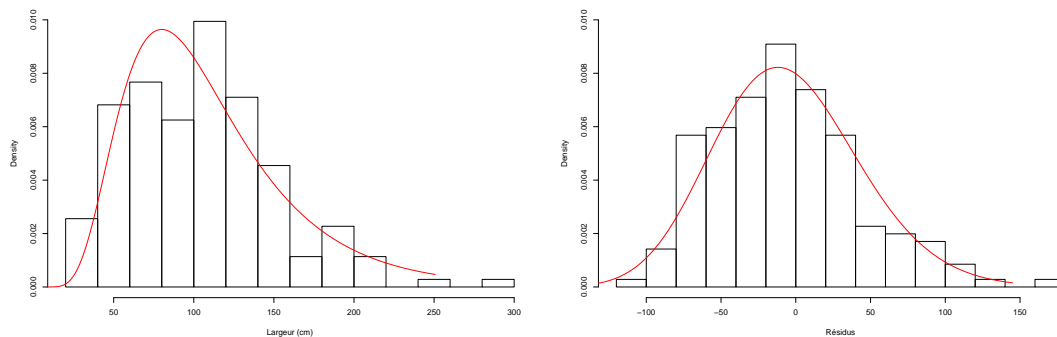


FIG. 3.16 – Histogrammes des largeurs observées par bief (cm) et des résidus de dérive amont-aval (cm)

**3.3.2.2.2 Test de corrélation spatiale du résidu** Afin de tester l'indépendance entre biefs successifs sur l'arbre  $G_P$ , on calcule pour chaque ordre de contiguïté amont-aval entre bief, un indice de Geary global (Cressie, 1993) (équivalent d'un variogramme normalisé). Afin de tester cette hypothèse d'indépendance de bief à bief, on compare ces valeurs d'indices de Geary à celles issues d'une distribution



sous hypothèse d'absence de corrélation spatiale (Fig. 3.17). Pour ce faire, on réalise des permutations aléatoires des valeurs de résidus par bief et à chaque permutation, on calcule les valeurs d'indices de Geary.

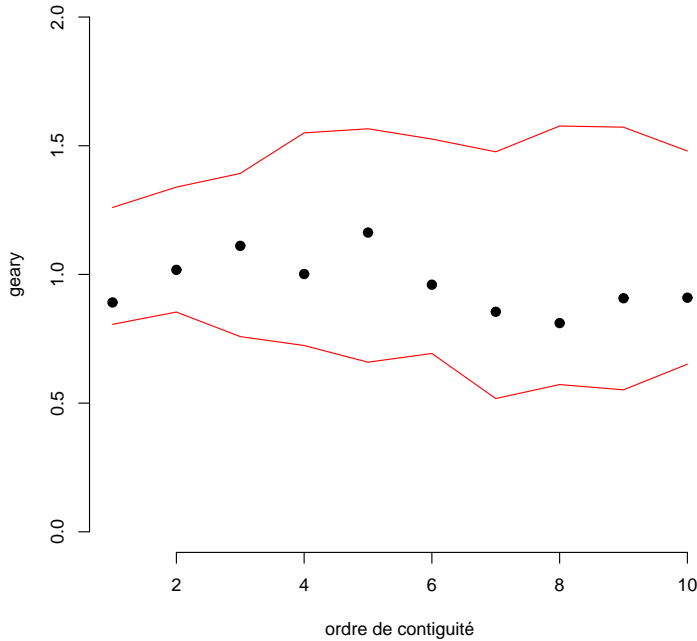


FIG. 3.17 – Evolution de l'indice de Geary par ordre de contiguïté amont-aval dans l'arbre  $G_P$  (ligne noire épaisse). Les courbes rouges délimitent une enveloppe à 95% de  $H_0$  correspondant à l'absence de corrélation spatiale.

Sur la figure 3.17, les indices de Geary observés sont bien compris dans l'enveloppe de  $H_0$ . On peut conclure sur une distribution des résidus indépendantes entre biefs (absence de corrélation de bief à bief).

### 3.3.2.2.3 Etude du processus de changement de bief

**Dérive Amont-Aval de l'intensité** Les intensités des processus  $\mathbf{X}$  et  $\mathbf{A}$  sont calculées localement suivant un noyau gaussien de 100 m d'écart-type avec les particularités de comptage introduites auparavant pour les extrémités de l'arbre et les confluences. Ces intensités locales de  $\mathbf{X}$ ,  $\hat{\theta}_X(s)$ , sont calculées pour un grand nombre de sites  $s$  de l'arbre et sont présentées en haut de la figure 3.18. Elles semblent a priori sans structure particulière.

Lorsque l'on projette les valeurs d'intensité locale  $\hat{\theta}_X(s)$  et  $\hat{\theta}_A(s)$  suivant un axe de position amont-aval  $x_2(s)$ , distance maximale à une source, on obtient le bas de

la figure 3.18. Cet indice de position dans l'arbre correspond à une pseudo-distance à la ligne de crête, distance reliée à l'unité morphologique et au type de parcellaire environnant, éléments qui conditionnent bien souvent le découpage du réseau en biefs. Selon ce nouvel indice de position amont-aval dans le réseau, on observe pour  $\mathbf{X}$ , comme pour  $\mathbf{A}$ , une intensité certes variable mais centrée sur une valeur constante, sans dérive apparente amont-aval. Seule une rivière "outlier" en amont présente des intensités très élevées (dues à une présence très particulière de buses tous les 5 m). A proximité de l'exutoire, on observe une intensité un peu plus faible mais peu significative. On considérera par la suite les processus  $\mathbf{X}$  et  $\mathbf{A}$  (et donc  $\mathbf{D}$ ) comme homogènes en intensité.

**Test de  $K(r)$**  Dans cette partie, on teste  $H_0$ , processus poissonnien homogène d'intensité constante  $\theta_X, \theta_A, \theta_D$  avec  $\theta_X = \theta_A + \theta_D$  (avec  $\theta_A$  et  $\theta_D$  homogènes). Pour ce, faire, on réalise un test de simulations sous  $H_0$  (simulations de  $\mathbf{A}$  additionné à  $\mathbf{D}$  fixé), puis on construit une enveloppe de confiance de  $K(r)$  sous  $H_0$ .

La fonction  $\hat{K}(r)$  observée (Fig. 3.19) est contenue dans l'enveloppe à 95 % de  $H_0$  pour toutes les distances explorées. Cependant, la limite de la fonction  $\hat{K}(r)$  observée est proche de la limite supérieure de l'enveloppe : cette situation est liée à la présence de canalisations enterrées (buses) qui créent des petits agrégats par paire sur le réseau.

On peut cependant conclure sur un processus ponctuel poissonnien homogène qui marque le changement de biefs sur le réseau de Roujan.

**3.3.2.2.4 Simulations de sections homogènes à distribution log-normale indépendante** Selon l'algorithme de simulation d'un processus de tessellation homogène dans un arbre orienté décrit auparavant, on simule un changement de bief par processus ponctuel poissonnien homogène puis on tire indépendamment pour chaque bief constitué dans la loi log-normale ajustée une valeur de résidu. On ajoute à ce résidu, une valeur de dérive  $m(u)$ , calculée suivant le critère  $x_3(s_\alpha)$ , avec  $s_\alpha$ , noeud amont du bief. On peut voir sur la figure 3.20 un exemple de largeurs simulées (bas) en comparaison avec les valeurs observées (haut).

**3.3.2.2.5 Liens entre paramètres du modèle poissonnien et paramètres du variogramme** Sur une simulation de largeurs par biefs sur l'arbre  $G_P$  suivant la méthode de tessellation exposée au dessus, on construit un variogramme expérimental à partir de couples de sites  $s_\alpha, s_\beta$  situés à une distance  $h$  sur des chemins amont-aval, selon la méthode et pour les sites présentés dans la section précédente sur les modèles géostatistiques. On reproduit cette estimation pour différentes simulations par biefs. Sur un grand nombre de simulations, on construit une enveloppe de confiance de variogrammes qui représente  $H_0$ , hypothèse de distribution de la variable largeur par biefs homogènes. Sur la figure 3.21, on représente cette enveloppe à 95 % construites avec 500 simulations et en gras, on replace le variogramme expérimental issu des données de largeur échantillonnées ponctuellement du jeu de données  $\mathbf{A1}$  et utilisé dans la section sur les modèles géostatistiques. On peut voir,

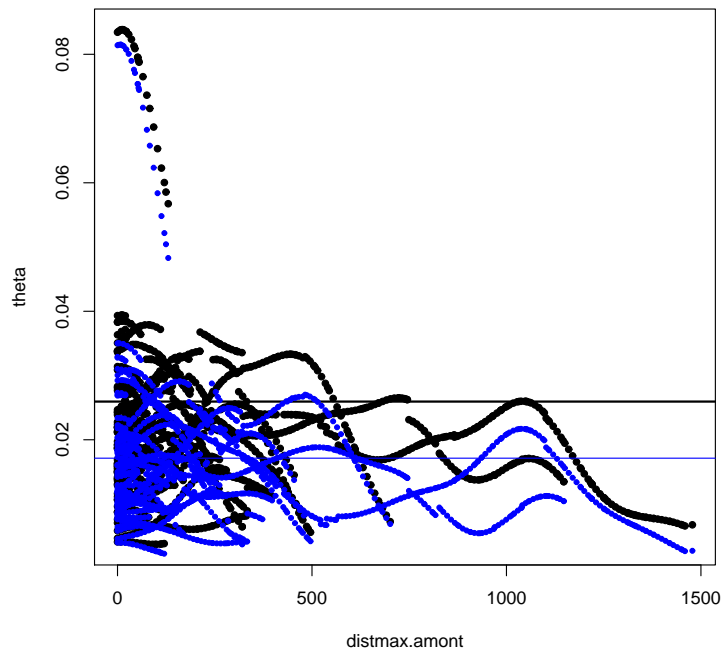
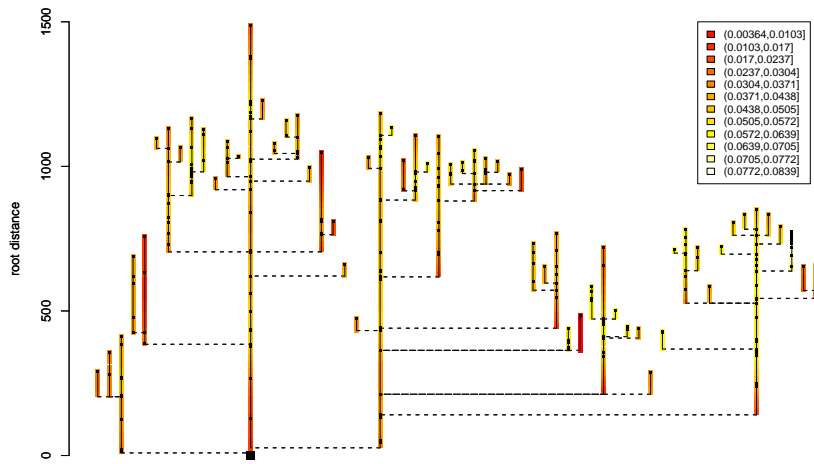


FIG. 3.18 – Intensité locale  $\hat{\theta}_X(s)$  dans l'arbre (Haut) puis projection des intensités  $\hat{\theta}_X(s)$  (noir),  $\hat{\theta}_A(s)$  (bleu) en fonction de la position amont-aval dans l'arbre  $x_2(s)$  (Bas). Les droites horizontales représentent les valeurs moyennes

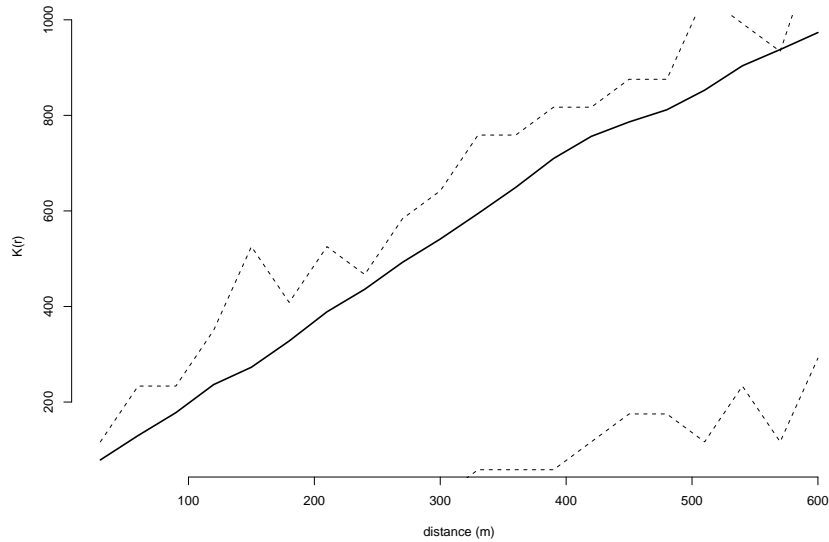


FIG. 3.19 – Test poissonnien de  $K(r)$

sur cet exemple, que le variogramme issues des données prises ponctuellement dans le jeu de données **A1** est bien au centre d’une hypothèse de processus de tessellation suivant un processus ponctuel poissonnien homogène de changement de bief et une distribution des largeurs par bief indépendante.

### 3.3.3 Discussion

Dans ce chapitre nous avons proposé une méthode de simulation spatiale de propriétés de réseaux arborescents et orientés qui simulent des variables réparties suivant des segments homogènes dans l’arbre appelées biefs. Ces méthodes s’appuient sur la combinaison 1- d’un processus ponctuel spatial aléatoire et homogène dans l’arbre qui segmente l’arbre en biefs et 2- d’un modèle de dérive amont-aval déterministe dans l’arbre puis 3- d’une loi de distribution par bief d’un résidu de dérive amont-aval sous hypothèse d’indépendance spatiale entre biefs.

Les modèles de processus ponctuels proposés ici sont uniquement poissonniens donc complètement aléatoires, ce qui rend l’application au support arborescent orienté possible. Des processus moins aléatoires (Neyman-Scott, Cox, Gibbs, ...) sur un tel support ne sont pas forcément possibles, ce qui limite l’adaptabilité de la méthode proposée à d’autres formes de structuration spatiale.

Toutefois, le modèle poissonnien proposé, associé à une dérive amont-aval, permet de simuler la largeur du réseau suivant une variabilité qui correspond à un modèle de variogramme exponentiel et dont les paramètres sont liés à l’intensité du processus poissonnien (pour la portée) et à la variance de la loi de distribution (pour le palier).

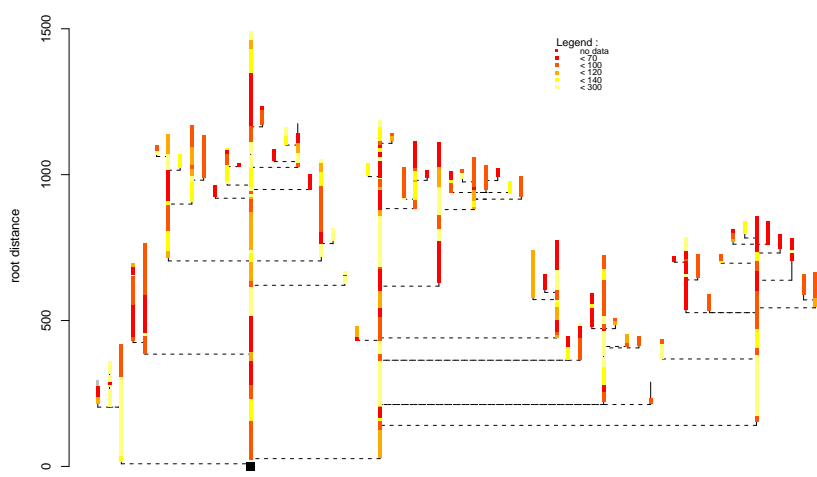
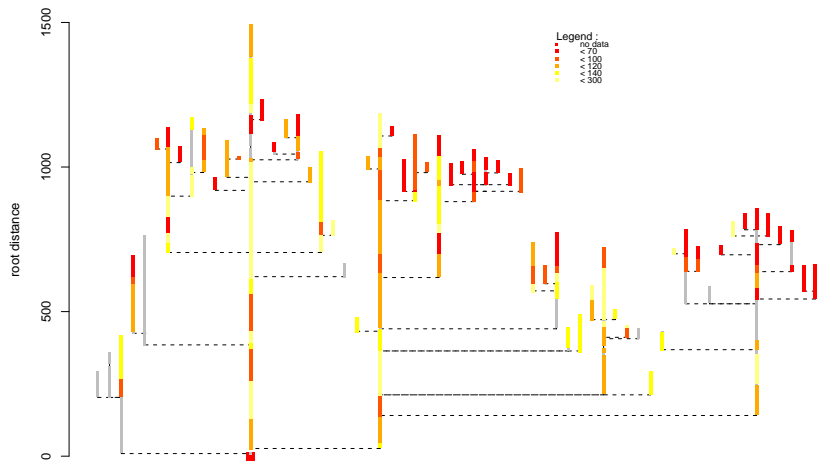


FIG. 3.20 – Valeurs observées par bief (Haut) et exemple d'une simulation par bief (Bas)

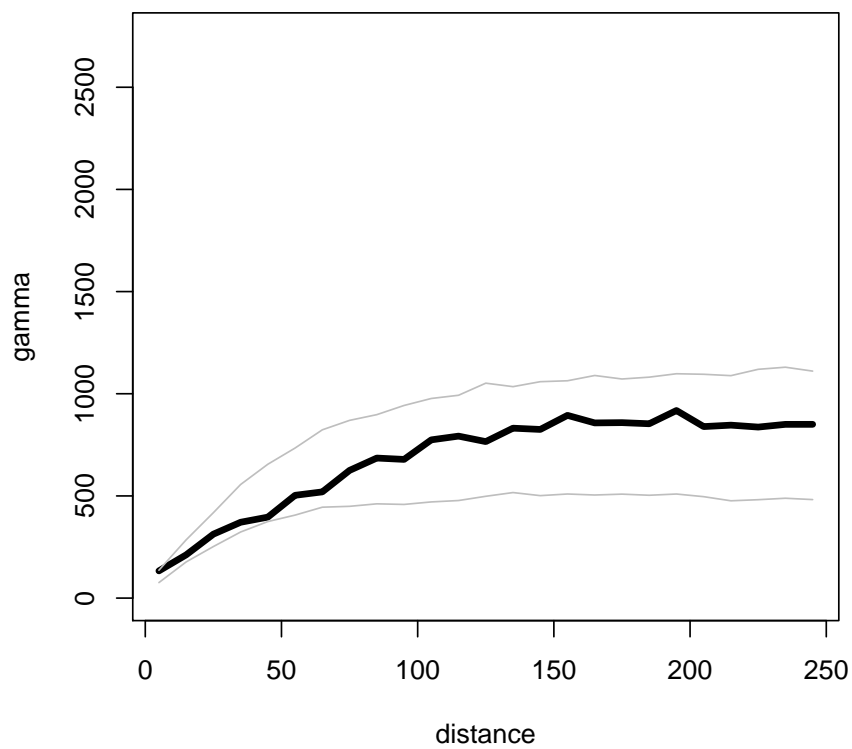


FIG. 3.21 – Variogramme expérimental amont-aval de la largeur du jeu de données **A1** (en gras) et enveloppes bilatérale à 95 % des variogrammes expérimentaux sur valeurs de largeur simulées par le modèle de tessellation  $1D$ .

## 3.4 Bilan

### 3.4.1 Conditions et limites d'application des modèles

Du point de vue méthodologique, avec les spécificités de modélisation que nous avons introduites, des méthodes de simulation de variabilités le long de réseaux arborescents orientés ont pu être proposées. Ces spécificités se résument 1– pour la méthode géostatistique, à l'introduction de probabilités conditionnelles dans le réseau qui permet ensuite de raisonner sur des droites orientées et 2– pour la méthode processus ponctuel, à la prise en compte d'effets de bordure et confluence spécifiques dans les estimateurs (intensité du processus, fonction de Ripley).

Les probabilités conditionnelles introduites dans le modèle géostatistique suppose une variable fortement conditionnée par l'orientation du réseau, i.e. les flux dans le réseau, comme la largeur. La méthode de simulation proposée n'est applicable qu'à des variables transformables en gaussiennes. On peut a priori l'étendre à des fonctions indicatrices (Emery, 2002). En revanche, les modèles de covariance admissibles sont nombreux sur cet espace (droites) mais les modèles géostatistiques proposés ne permettent pas pour autant de faire de l'estimation (krigeage). Sur ce point, des travaux méthodologiques récents aux objectifs similaires sont apparus dans la littérature (VerHoef *et al.*, 2006; Cressie *et al.*, 2006). Ces travaux utilisent les propriétés des moyennes mobiles par noyau appliquées à des mouvements browniens le long du réseau (bruit blanc) qui conduisent à des covariances de structure connues, admissibles sur des réseaux (VerHoef *et al.*, 2004), et qui permettent de faire de l'estimation par krigeage.

La méthode processus ponctuel est fortement guidée par la prise en compte d'un caractère artificiel a priori sur les variabilités étudiées. La simplicité du processus de changement de bief, poissonnier et homogène, somme de deux sous-processus indépendants, eux aussi poissonniers et homogènes et dont un des sous-processus est fixe car conditionné par l'architecture du réseau, rend la méthode de simulation du processus quasi immédiate. Les processus étant complètement aléatoires (poissonniers), les rivières sont de fait indépendantes les unes des autres et l'arbre peut être traité dans ce cas comme un ensemble de segments de droite indépendants. Tout autre processus ponctuel moins "indépendant" serait impossible. Pour ce qui est de la distribution de la largeur par bief, l'indépendance entre biefs, à un modèle de dérive près, rend aussi la simulation de la distribution des largeurs aisée. Si ce n'avait été le cas, la dépendance selon ce support discret serait également à modéliser selon une approche similaire à celle développée dans la partie géostatistique.

### 3.4.2 Validation

#### 3.4.2.1 Paramètres des modèles

Les méthodes proposées n'ont jamais été croisées sur un autre jeu de données que celui qui a servi à l'inférence des modèles. Un seul réseau a été étudié. On a donc a priori aucune idée de leur degré de généralisation. Un moyen d'effectuer cette validation croisée serait, sur un ou d'autres réseaux, de tester la stabilité des structures de modélisation (forme fonctionnelle des dérives, des variogrammes) et des paramètres estimés. Pour les simulations, un seul jeu de paramètres a été utilisé. N'ayant aucune idée a priori de la distribution de ces paramètres, nous n'avons pas cherché à les rendre aléatoires. Là encore avec d'autres réseaux étudiés, d'autres

estimations de modèles, des simulations plus objectives sur les incertitudes prenant en compte une variabilité des paramètres pourraient être mises en oeuvre.

### 3.4.2.2 Réalisme des simulations

Finalement, pour la largeur du réseau de Roujan, quelle méthode donne les simulations les plus réalistes ? Les données pour ces deux méthodes ont été acquises à trois ans d'intervalle. La variable étudiée (largeur) est suffisamment statique pour pouvoir croiser les résultats. Prenons un exemple d'une rivière du réseau de Roujan (la numéro 30), qui représente 5 % du linéaire du réseau, afin d'explorer qualitativement cette question. Sur la figure 3.22, on représente à droite les vraies données **A1** (en noir) et en bas, en rouge, une simulation non conditionnelle géostatistique. A droite, on représente les vraies données **A2** (en noir) et en bas, en rouge, une simulation par processus ponctuel. Ces deux méthodes reproduisent des structures de covariance exponentielle. Sur cet exemple, sous hypothèse d'une stabilité de cette variable sur la durée séparant les jeux de données, la méthode de simulation géostatistique semble donner les simulations les plus "réalistes".

### 3.4.3 Extensions possibles de ces approches de simulation

Sous réserve de la mise en place de tests de réalisme des simulations, la simulation de variabilités par segments homogènes indépendants pourrait être envisagée pour d'autres variables d'intérêt de ces réseaux (même processus de changement de bief), notamment celles liées à l'occupation du sol environnante du réseau : l'enherbement ou autres variables explicatives du coefficient de frottement hydraulique. L'approche géostatistique proposée telle quelle, au contraire, n'est pas indiquée pour de telles variables. En revanche, par extension multivariée, cette méthode pourrait également être utilisée pour simuler conjointement largeur et profondeur du réseau, non indépendantes entre elles, soit le gabarit complet du réseau. Cette extension multivariée ne pose pas à priori de difficultés sous hypothèse d'indépendance conditionnelle conjointe entre largeur et profondeur entre branches parallèles.

### 3.4.4 Simulations ou simulations conditionnelles ?

Dans ce chapitre, nous n'avons présenté que des résultats de simulation sur tout le réseau, le terme conditionnement étant relatif à la structure du réseau qui conditionne, soit un ordre de simulation des rivières (géostatistique), soit une partie du découpage en biefs. Pour la partie géostatistique, l'algorithme de simulation proposé reste cependant valable pour des simulations conditionnées par quelques points de données sur le réseau. Pour la méthode processus ponctuel, des méthodes de simulation conditionnelles d'un processus ponctuel en  $1D$  existent déjà (Lantuejoul, 2002, P.124), et pourraient être appliquées à notre cas complètement poissonnien.

Pour la partie géostatistique, on peut illustrer cette différence en prenant le cas de la rivière 30 du réseau de Roujan utilisant les paramètres des modèles ajustés en section 1 (dérive, variogramme, distribution des résidus). Sur la figure 3.23, on représente, en haut les vraies données sur cette rivière 30 puis une simulation non conditionnée et conditionnée par 8 points de données pris au hasard sur cette rivière (points gris en gras). Les lignes verticales en pointillé désignent l'emplacement d'une confluence.



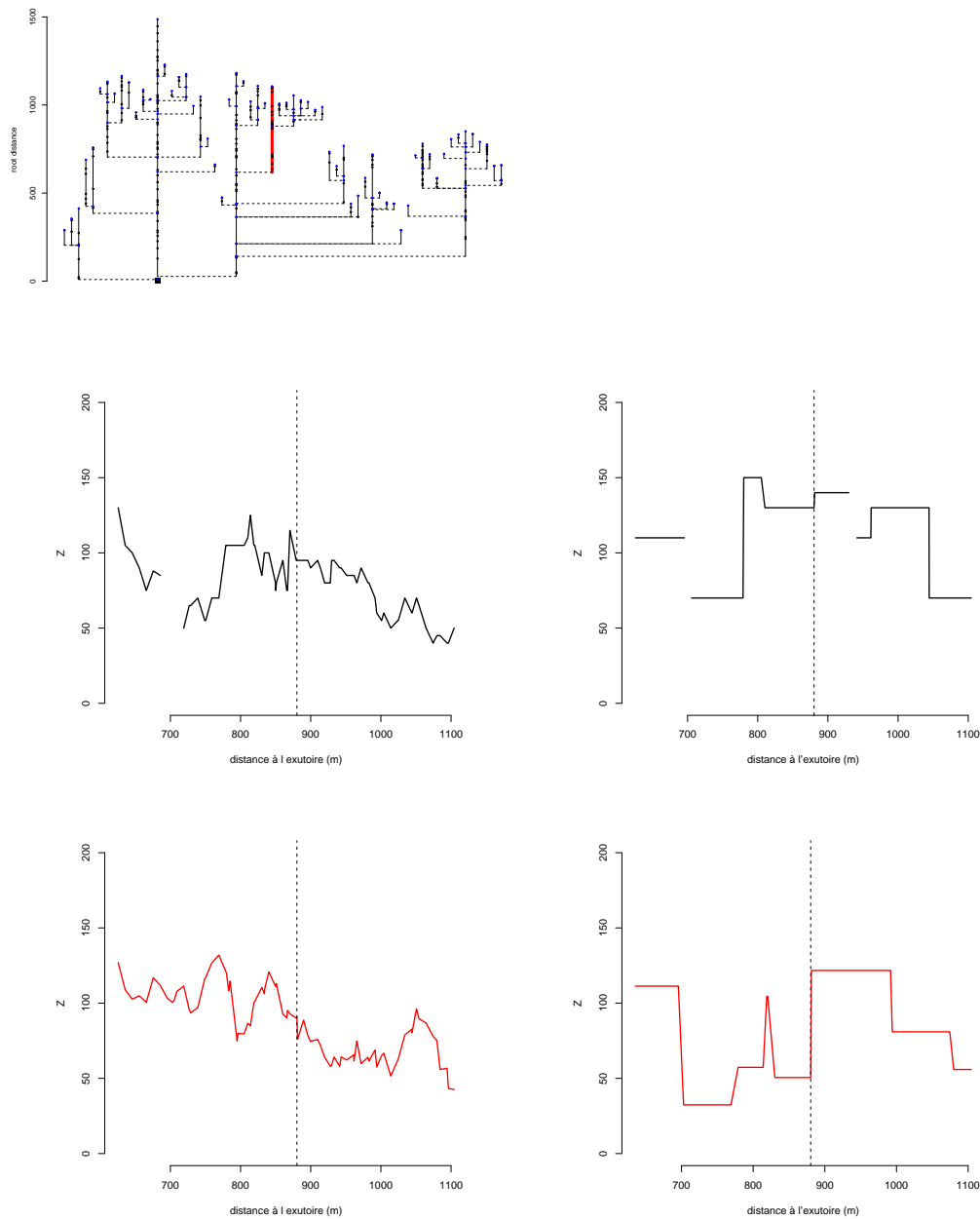


FIG. 3.22 – Largeurs observées (en noir) et simulées (en rouge) sur la rivière 30 (en trait rouge épais sur le dessin du réseau en haut) ; Données **A1** et simulation géostatistique (Gauche) ; Données **A2** et simulation processus ponctuel (Droite) ; les lignes verticales en pointillé marquent une confluence.

Le conditionnement par quelques points de données (points gris) améliore le réalisme des simulations. Cela signifie que pour un autre réseau, si les paramètres

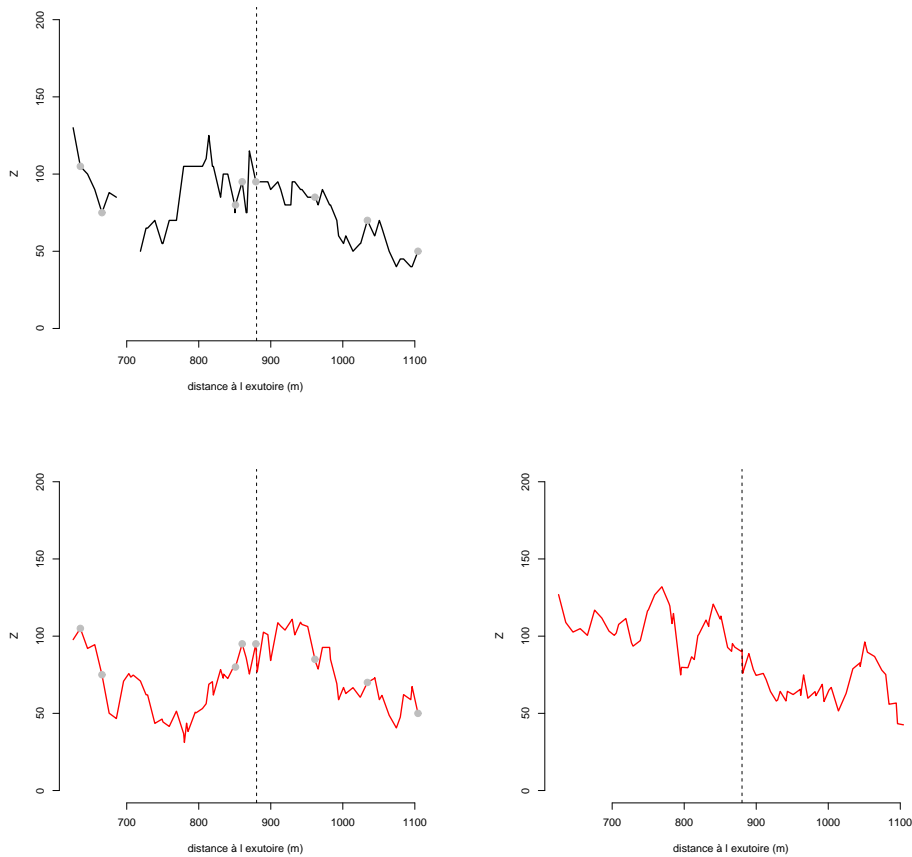


FIG. 3.23 – Largeurs ( $Z$ ) observées sur le jeu de données **A1** (en noir) et simulées (en rouge) sur la rivière 30 : non conditionnée par des données (droite) et conditionnée (gauche)

des modèles (dérive, variogramme) sont les mêmes que pour celui de Roujan, quelques observations de largeur sur ce réseau peuvent être introduites dans les simulations de largeur pour maximiser leur réalisme.

### 3.4.5 Complémentarité et additivité des modèles

Les résultats de comparaison des simulations suivant les deux approches développées n'ont rien de surprenant. On sait que plusieurs modèles et algorithmes peuvent donner les mêmes structures spatiales en terme de covariance et les mêmes distributions (Chiles et Delfiner, 1999, p.233). De ce fait, sur le cas d'étude, des segments homogènes et indépendants reproduisent une structure de covariance fonction de la distance amont-aval dans le réseau très marquée. Ce qui peut paraître plus surprenant est qu'il existe une variabilité locale de la largeur du réseau assez forte, pas vraiment distribuée selon des sections homogènes. Pourtant la portée modèle de variogramme correspond bien à l'ordre de grandeur de la longueur de bief moyenne

sur le réseau (découpé en biefs selon la méthode AQUAE (Lagacherie *et al.*, 2004)). On peut penser que la largeur du réseau est une variable dont le processus original est bien artificiel selon des sections homogènes mais avec le temps (le réseau de Roujan est ancien), le naturel "reprend le dessus" (érosions latérales lors des crues, effondrements, dépôts...). En revanche, quelque soit l'approche, on voit, sur cet exemple de largeur de réseau anthropique, que contrairement à quelques idées reçues, les petits ruisseaux ne font pas les "grandes rivières" (processus naturel), les dérives amont-aval étant assez faibles, et que le réseau n'est pas non plus de dimensions constantes (liés par exemple aux dimensions de pelles mécaniques), résultat de l'action de l'homme uniquement (processus anthropique).

Un modèle plus général, adapté à des variabilités semi-anthropiques ou semi-naturelles sur ces réseaux, serait un mélange paramétré de ces deux approches de modélisation : un processus ponctuel qui marque des changements d'espérance constante par bief associé à un variogramme des résidus autour de cette espérance.



FIG. 3.24 – Extrait de (Ahern, 2006)



# Chapitre 4

## Simulations conditionnelles d'arbres orientés dans un lattice orienté

### Sommaire

---

4.1	Introduction . . . . .	95
4.2	Conditionnement : Télédétection de la présence de fossés sur un lattice de limites cadastrales . . . . .	96
4.2.1	Détection de la présence de fossés sur profils altimétriques LiDAR . . . . .	96
4.2.1.1	Introduction . . . . .	96
4.2.1.2	Materials and methods . . . . .	96
4.2.1.2.1	Study zone . . . . .	97
4.2.1.2.2	Field data . . . . .	98
4.2.1.2.3	LiDAR data . . . . .	98
4.2.1.2.4	Methods . . . . .	98
	Step 1 : elevation profiles estimation . . . . .	99
	Step 2 : curve shape analysis . . . . .	101
	Step 3 : classification method . . . . .	105
4.2.1.3	Results and discussion . . . . .	106
4.2.1.3.1	Elevation profile interpretation . . . . .	106
4.2.1.3.2	Automated profiles classification . . . . .	108
	Classification accuracies computation . . . . .	108
	Classification dimension reduction . . . . .	108
	CART Classification accuracies . . . . .	109
	Predicted classes mapping . . . . .	110
	Sensitivity analysis . . . . .	111
4.2.1.4	Discussion . . . . .	113
4.2.2	Classification binaire des arêtes du lattice cadastral à partir des profils LiDAR classés . . . . .	115
4.2.2.1	Introduction . . . . .	115
4.2.2.2	Données . . . . .	115
4.2.2.3	Méthodes . . . . .	116

4.2.2.3.1	Choix des variables descriptives par arête	116
4.2.2.3.2	Arbre de classification	116
4.2.2.4	Résultats	117
4.2.2.5	Conclusion	120
<b>4.3</b>	<b>Algorithme de simulation conditionnelle d'un réseau de fossés dans le lattice des limites cadastrales</b>	<b>121</b>
4.3.1	Introduction	121
4.3.2	Bibliographie : Simulations d'arborescences	122
4.3.2.1	Indices descriptifs de l'architecture d'arborescences	122
4.3.2.1.1	Topologie d'une arborescence	122
	Taxonomie de Strahler	122
	Distribution des sources	123
	Matrice de ramification	123
4.3.2.1.2	Géométrie d'une arborescence	124
4.3.2.2	Lois et simulations non-conditionnelles d'arborescences	125
4.3.2.2.1	Approches fractales	125
4.3.2.2.2	Approches basées sur les matrices de ramification	127
4.3.2.3	Simulations conditionnelles de réseaux	127
4.3.2.4	Bilan	128
4.3.3	Méthode de simulation conditionnelle d'arbres orientés dans un lattice orienté	129
4.3.3.1	Schéma général	129
4.3.3.2	Étape 1 : Densifier le lattice par reconnection des îlots parcellaires	130
4.3.3.3	Étape 2 : Former un "tronc" orienté à partir des arêtes données	131
4.3.3.4	Étape 3 : Élagage et branchements sur le tronc pour converger vers une distribution de sources	134
4.3.3.4.1	Outils	135
4.3.4	Cas d'étude	135
4.3.4.1	Site d'étude test	135
4.3.4.2	Topologie du vrai réseau	135
4.3.4.3	Données pour l'orientation du Lattice	137
4.3.5	Résultats	138
4.3.5.1	Ajout d'arêtes entre îlots parcellaires	138
4.3.5.2	Orientations du lattice	138
4.3.5.2.1	Sensibilité à $\Delta Z$	138
4.3.5.2.2	Degrés inférieurs des noeuds du lattice orienté	140
4.3.5.3	Distribution spatiale des sources	141
4.3.5.4	Distribution d'un ensemble de simulations	142
4.3.5.4.1	Critères topologiques	143
4.3.5.4.2	Critères géométriques	144
4.3.5.4.3	Critères topographiques	146
4.3.6	En conclusion	147
<b>4.4</b>	<b>Bilan</b>	<b>148</b>

---

## 4.1 Introduction

---

*Après s'être intéressé au développement de méthodes pour simuler des variabilités spatiales le long d'une arborescence, on s'intéresse dans ce chapitre à la simulation de l'arborescence elle-même. Compte-tenu des hypothèses de travail, l'objectif de ce chapitre est donc de proposer une méthode de simulation qui puisse s'appliquer à l'architecture d'un réseau hydrographique artificiel d'un paysage cultivé tout en intégrant des données partielles sur ces réseaux. Pour ce faire, les simulations sont réalisées dans l'espace des limites cadastrales, dit lattice parcellaire ou cadastral, et elles sont conditionnées par l'utilisation de données de télédétection LiDAR. Ces données sont envisagées pour 1- détecter autant que faire se peut, les fossés dans le lattice parcellaire et 2- orienter une simulation de réseau arborescent de l'amont vers l'aval qui connecte les morceaux de réseau détectés.*

*Ce chapitre est organisé en trois sections. La première correspond à une approche métrologique des réseaux de fossés par télédétection. Elle s'intéresse notamment aux possibilités de détection des fossés par télédétection LiDAR. Elle est elle-même subdivisée en deux sous-sections, relatives à deux échelles spatiales d'étude. La première est en anglais ; elle est extraite d'une publication acceptée avec révisions mineures dans la revue *International Journal of Remote Sensing* et intitulée "Agrarian landscapes linear features detection from LiDAR elevation profiles : application to artificial drainage network detection" (Bailly et al., in press). Elle étudie les possibilités de détection automatique de la présence de fossés sur des profils altimétriques LiDAR situés le long des limites cadastrales (étude réalisée sur le site test **B**). La deuxième sous-section est une homogénéisation (on parlera ensuite de "filtrage") par arête du lattice parcellaire des résultats de détection obtenus précédemment. Elle aboutit à une classification en mode binaire (présence-absence de fossés) de chaque arête du lattice cadastral (utilisation des sites tests **B** et **C**).*

*La deuxième section porte sur la simulation conditionnelle de réseaux arborescents en partant des résultats de détection homogénéisés précédents. La méthode mise en oeuvre s'appuie sur un algorithme de simulation stochastique, orchestrant les différentes contraintes topographiques, topologiques et le conditionnement par les portions de réseaux détectées (utilisation du site test **D**). Le réalisme de l'architecture des résultats de simulation est évalué suivant des critères topologiques et géométriques issus de la littérature.*

*La dernière section fait un bilan des méthodes de simulations conditionnelles de réseaux artificiels de paysages cultivés mises en oeuvre .*

---



## 4.2 Conditionnement : Télédétection de la présence de fossés sur un lattice de limites cadastrales

### 4.2.1 Détection de la présence de fossés sur profils altimétriques LiDAR

Extrait de "*Agrarian landscapes linear features detection from LiDAR elevation profiles : application to artificial drainage network detection*" (Bailly *et al.*, in press).

#### 4.2.1.1 Introduction

As for most surface mapping problems, remote sensing can be a candidate for agrarian landscapes' linear elements<sup>1</sup> mapping. But when considering these element's dimensions, i.e. width or height, only very high spatial resolution remote sensing data will suit. Linear features detection from remote sensing data is usually performed on images using morphological filters (Tripathi *et al.*, 2000), pixel classifications (Thornton *et al.*, 2006), segmentations (Mason *et al.*, 2003) or drainage algorithms when the regular grid is a digital terrain model (D.T.M) (Tarboton, 1997). For the latter, the few studies on artificial drainage network detection in agrarian landscapes showed that drainage algorithms on D.T.M are not suitable in the case of anthropogenic networks (Comoretto, 2003). Up to now, no generic methodology has been proposed for mapping agrarian landscapes' linear elements. The most probable reason for this is that data on elevation had too low resolution. Now, LiDAR reveals new data that looks promising for such goals. LiDAR data have been already applied to linear element mapping, e.g. : erosion gullies (Ritchie *et al.*, 1994), dams (Pereira et Wicherson, 1999) or flood plain hedgerows (Marks et Bates, 2000). However, these studies performed visual characterization of LiDAR data in predetermined areas for wide linear elements (from 3 to 40 meters) and they did not propose any automated methodology.

In this paper, we propose a methodology that aims to automatically detect or characterize agrarian landscapes' linear features from LiDAR elevation data. This methodology includes three successive steps : 1- building elevation profiles on a set of pre-located sites from LiDAR raw data, 2- deriving profile concavity indicators by means of two curve shape analysis methods, namely wavelet transforms and a watershed algorithm (Soille, 1999), to build synthetic profile discontinuity descriptors, 3- classifying the profiles with the concavity indicators using the CART method (Breiman *et al.*, 1984).

The methodology we developed was applied to detect a ditch network, within a 2 km<sup>2</sup> study zone located in the Peyne catchment in the south of France. This zone is typical of the Mediterranean vineyard landscapes where pesticide transfers within ditch network have been observed for many years (Louchart *et al.*, 2001). This paper is organized as follows. Section 4.2.1.2 presents first the studied agrarian linear features, i.e. ditches, the data and methods we used, with a focus on 1D wavelet transforms. In section 4.2.1.3, we present and discuss step by step the results. Sensitivity analysis regarding inhomogeneity in space of the LiDAR data are shown at end of the section 4.2.1.3.

#### 4.2.1.2 Materials and methods

---

<sup>1</sup>hedgerows, ditches and bench terraces

**4.2.1.2.1 Study zone** The study zone is a 2 km<sup>2</sup> area located in the Peyne catchment in the Languedoc (South of France) (Fig. 4.1). This area was selected as representative of the Languedocian vineyard landscape. The area includes plateaus, sloping areas with terraces (made to limit soil erosion), gentle foot slopes made of alluvial and colluvial deposits and a flat depression, successively, from upstream to downstream (Fig. 4.1).

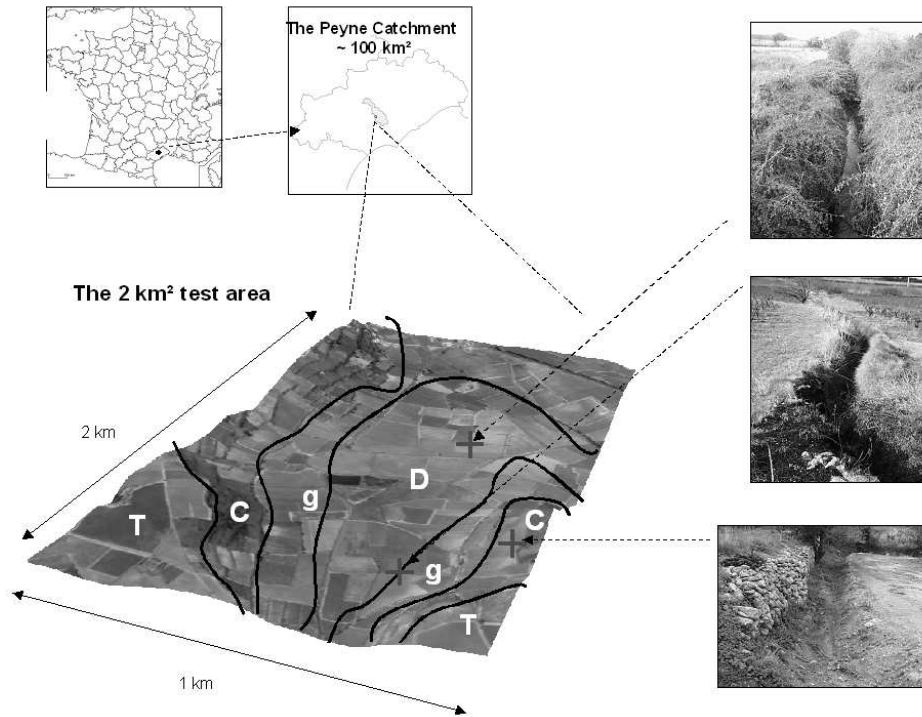


FIG. 4.1 – The 2 km<sup>2</sup> study zone is depicted down left in 3D on images with an elevation factor equal to 3. Morphological units are delineated in black lines : T (plateaus), D (depression), C (areas with high slopes and terrassettes), g (glacis and alluvial terraces) ; 3 examples of ditches photographs are on the right

This zone is equipped for pollution and flooding measurements (Lagacherie *et al.*, 2004). It contains a complex and dense artificial drainage network 23 km long that drains land plots and prevents erosion in case of huge runoff. The landscape positioning of the ditches that mainly compose this network is highly variable and depend closely on ditch functions. When the ditch function is to capture surface waters, in the case of high slopes, they are perpendicular to terrain slope, especially at the step foot of terrassettes. When function is water transport, they follow slope directions. When ditch function is to drain subsurface waters, there is no evident relationship with local relief. Ditch geometrical sections are also highly variable with areas ranging from 0.06 to 5 square meters and widths ranging from 0.5 to 2.5 meters. Furthermore, highly variable maintenance of ditches by farmers causes a high variability of vegetation inside and over ditches (Fig. 4.1).

**4.2.1.2.2 Field data** A geographical database on the ditch network was already available for the study area (Lagacherie *et al.*, 2004). Ditches were digitized using the lattice of plot boundaries as a background layer. Consequently, the location of each ditch match those of a field plot boundary. At the same time as the LiDAR survey, additional ground truth observations on the ditches were acquired. These ground truth observations, located on 99 ditch sites, included transversal topographic profiles and qualitative attributes describing vegetation presence over ditches.

**4.2.1.2.3 LiDAR data** The LiDAR first and last pulse scatter of points were acquired on the study area in June 2002. The Falcon II Toposys LiDAR system (*www.toposys.com*) was fixed on a helicopter surveying at an elevation of 900 m and with an 83 MHz emission rate for the laser pulses.

*Voir en annexe p. 174 la présentation du principe du LiDAR*

This survey configuration was in a swing mode (Wiechert *et al.*, 2000), providing Near Infra Red (NIR) first and last <sup>2</sup> pulses corresponding to topographic X,Y and Z point values. LiDAR NIR pulse intensity was not registered in this configuration mode. This LiDAR system configuration gave scatters of 3D points with a mean density of 10 points per square meter. These points are centers of laser beam footprints on terrain or surface, irregularly distributed in space. These footprints can be considered disks 60 cm in diameter. Laser beam footprints overlap greatly along track direction as seen on the left of figure 2 where track direction is about NE-SW. Topographic LiDAR points are then highly spatially correlated and provide no abrupt change in height (Wiechert *et al.*, 2000). Post survey accuracy controls on LiDAR last-pulses topographic points on flat and non vegetated areas exhibit a gaussian noise with a standard deviation of about 0.06 m for altimetry (Saidouni, 2003).

Figure 4.2-A maps two contrasting examples of LiDAR points sampling design around a field plot boundary corresponding to the ditches and their consequences on terrain modelling (Fig. 4.2-B). The upper example is a case where 1–there is little vegetation inside and around the ditch at survey time; consequently, vegetation does not mask the ditch concavity; and 2– LiDAR points are evenly distributed across the ditch, that is, directed perpendicular to the flight track. However, laser pulses are not powerful enough to penetrate vine row vegetation as shown on the elevation profile centered on the ditch where no terrain point is registered below a vine row. The lower example is a case where 1– LiDAR points are irregularly distributed (see gaps in the elevation profile) because the ditch is directed in parallel to the flight track and 2– laser pulses do not penetrate the hedgerow and mask ditch elevation concavity (Fig. 4.2-B).

**4.2.1.2.4 Methods** The proposed approach is based on the underlying hypothesis that 1) ditches are exclusively located at field boundaries (this was observed in our case study), and that 2) ditches are expected to produce a concavity within an elevation profile directed across the ditch (see figure 4.2-B). Consequently,

---

<sup>2</sup>First or last significative (then thresholded) pulses of light energy registered in the LiDAR reception system

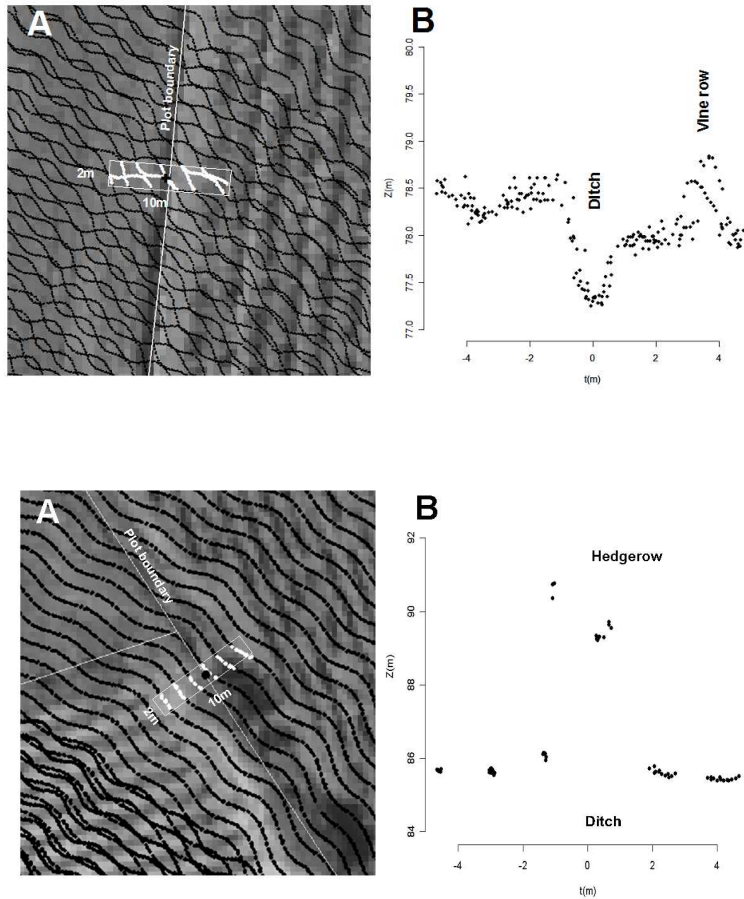


FIG. 4.2 – A : 2 examples of LiDAR points spatial sampling design around ditches - LiDAR points are depicted as black points on an panchromatic 0.5 m resolution image. Selected LiDAR points within a 2 by 10 m windows centered on a ditch on a plot boundary are depicted in white then B : plotted in elevation profiles

the approach includes three successive steps that involves three distinct procedures. The first procedure uses LiDAR data to estimate elevation profiles centered on field boundaries. The second procedure is a curve shaped analysis that derives concavity indicators from the elevation profiles. The third procedure is a boolean classification procedure that discriminates ditches from " non ditches " on the basis of previously calculated indicators.

**Step 1 : elevation profiles estimation** According to the hypothesis presented before, elevation profile estimations from the scatter of LiDAR last pulse points are computed on field boundary sites. In essence, this reduces the space for ditch detection to pre-located sites on this lattice. We automatically estimated elevation profiles perpendicular to and centered on field plot boundaries with the following

four step process :

1. Sites along plot boundaries were regularly selected with a 10 meters lag; site coordinates and boundary azimuth on the sites were automatically calculated from the plot boundary database.
2. From those coordinates and azimuth, LiDAR last pulse points were selected for each site within a 10 by 2 meters window centered on the site and perpendicular to the plot boundary as seen in the example in figure 4.2-A ; 2 meters was chosen because it was the smallest width that permitted us to cover two parallel sample lines<sup>3</sup> in the across-track direction. The 10 meter profile was considered long enough to detect elevation discontinuities from ditches with different profile trends.
3. From each set of selected LiDAR points (Fig. 4.2-B), we calculated a regular and dyadic profile with 32 bins. This was done by computing the minimum elevation envelope from the selected points, seen as the wide grey line on the figure 4.3. In the case of gaps in LiDAR sampling as seen in the lower example in the figure 4.2-B, linear interpolation was performed.
4. We added a translation process to homogenize the location of the central concavities between the elevation profiles, these locations being expected to be altered by planimetric positioning errors in the LiDAR data and field plot digitized coverage. The profiles are translated in order to force a local minimum into the 17th bin as seen in figure 4.3. This was done with a maximum translation lag (1 m which includes most planimetric positioning deviations) to always stay on a local and central minimum.

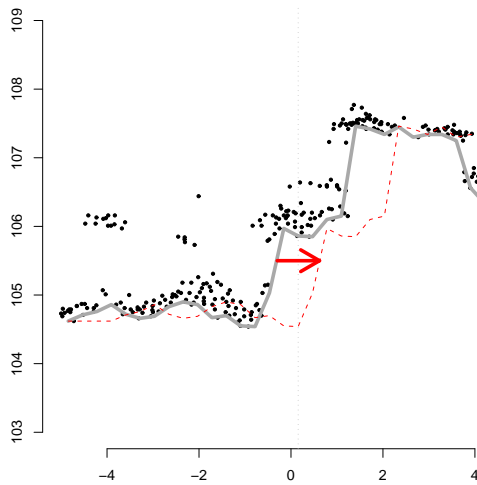


FIG. 4.3 – Regular profile estimation from LiDAR points and profile translation process

---

<sup>3</sup>coming from two adjacent optical fibers in the scanning LiDAR system

This process was validated by comparison between estimated profiles and ground truth profiles (Saidouni, 2003). 6461 profiles along plot boundaries were computed on the  $2km^2$  study zone. For each profile, exhaustive ground survey (Lagacherie *et al.*, 2004) informed us if it was centered on a ditch or not. We called further the set of 2531 of 6461 profiles that were centered on ditches, the *D* profiles or sites. The other 3930 profiles, not centered on ditches, were called the *ND* profiles or sites.

**Step 2 : curve shape analysis** As ditches are expected to produce concavity within an elevation profile, curve shape analysis methods were applied to detect concavities that were expected to have various shapes and to be located in various relief contexts but always located in the middle of the profile. Several methods can be candidates for such problems, coming from signal processing and morphological signal analysis. The utility of wavelet transform (Antoine *et al.*, 1997) has been widely proved in various curve classification problems : in forest stand analysis from canopy LiDAR topographic profiles (Ollier *et al.*, 2003), in time series processing (Struzik, 2000) and in LiDAR backscattering analysis (Wang et Sassen, 2002) for discontinuity detection. Wavelets are transforms that allow analysis of part of the noise-free signal or to select the resolutions and the signal locations consistent with the studied shapes. In essence, wavelet transform reduces dimensions for curve classification (Rouvière, 2005).

Another popular way to detect and characterize concavities in a curve is the use of morphological functions (Soille, 1999). The difference between the closing of the signal and the signal itself can perform concavity detection but only with concavity dimensions close to structural element size. A more general way to detect and characterize concavities is the use of a watershed transform that segments concavities into a signal with a flooding algorithm.

We chose here to compare two methods : decimated Haar wavelet transforms and a specific watershed algorithm. Both methods are suitable for central concavity detection. Moreover, with wavelet transforms, we can keep information on the general trend of the signal that traduces the morphological context (slope, curvature, terrasse step, ...). The watershed algorithm is strictly dedicated for central concavity detection.

### *Wavelet transform*

### *Wavelet properties*

1D wavelet transform is a transform of a 1D signal  $z$  with a pair of orthogonal wavelet functions :  $\phi$ , the father function and  $\psi$ , the mother function. The wavelets pair presented in figure 4.4 is the first orthogonal wavelet functions pair proposed by Haar (Haar, 1910) that presents simple shapes, easily interpretable and reliable for ditches concavity detection in LiDAR profiles. Moreover, the simple shapes and compactness of the Haar wavelet functions make them suitable for the short 1D signal wavelet transform.

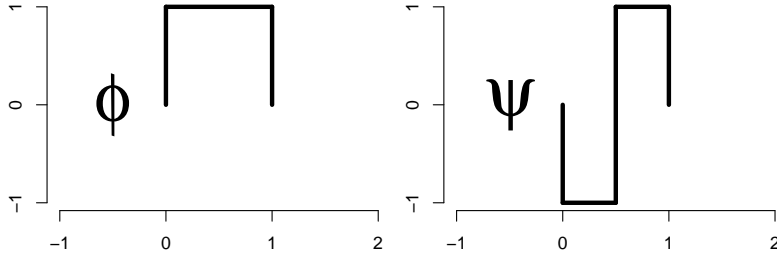


FIG. 4.4 – Haar wavelet pair : father wavelet  $\phi$  on the left and mother wavelet  $\psi$  on the right

The wavelet transform of a 1D continuous signal  $z(t)$  is given by :

$$z(t) \approx \sum_k S_{J,k} \phi_{J,k}(t) + \sum_{j=1}^{J-1} \sum_{k=0}^{n_j-1} d_{j,k} \psi_{j,k}(t) \quad (4.1)$$

In equation 4.1,  $j$  is called *resolution* or *scale parameter* going from  $J$ , the coarser resolution, to 1, the finest resolution.  $k$  is a *translation parameter* going from 0 to  $n_j - 1$ , with  $n_j$  corresponding to the number of coefficients at resolution  $j$ .

$\int \phi(t) dt = 1$  then  $\phi$  approximates the signal trend.

$\int \psi(t) dt = 0$  then  $\psi$  is used to approximate the signal details.

The principle of the wavelet transform is to approximate  $z$  with decreasing resolution through scaling and translations of  $\psi$  and  $\phi$  functions. The coefficients  $S_{J,k}, d_{J,k}, \dots, d_{1,k}$  are the wavelet transform coefficients. The  $S_{J,k}$  are the trend coefficients and  $d_{j,k}$  are the detailed coefficients.

In case of a discrete signal  $z_1, \dots, z_n$ , we can use the decimated wavelet transform (D.W.T). D.W.T calculates the coefficients of the wavelet approximation using a fast pyramid algorithm (Mallat, 1989). We can find an exact definition for D.W.T coefficient calculations in (Bouchereau, 1997, p.59).

Voir en annexe p. 179 une illustration de la transformée pas à pas en ondelettes discrètes (D.W.T) 1D

Figure 4.5 shows 4 virtual profiles D.W.T. These 4 profiles are typical profiles with central sharp concavities corresponding to ditches. For each one of the four cases, the virtual profile is the upper wide black curve. According D.W.T coefficients on 4 resolutions are plotted below : from top to bottom, trend coefficients  $S_{4,k}$  values are plotted in grey lines, then detailed coefficients  $d_{j,k}$  from coarser scale ( $j = 4$ ) to the finest ( $j = 1$ ) are plotted in black lines. The white thin lower curve is the reconstructed profile using the D.W.T coefficients and the Mallat algorithm (Mallat, 1989) for equation 4.1. This reconstructed profile, shown in white, is on the black wide curve corresponding to the genuine profile.

These figures show that :

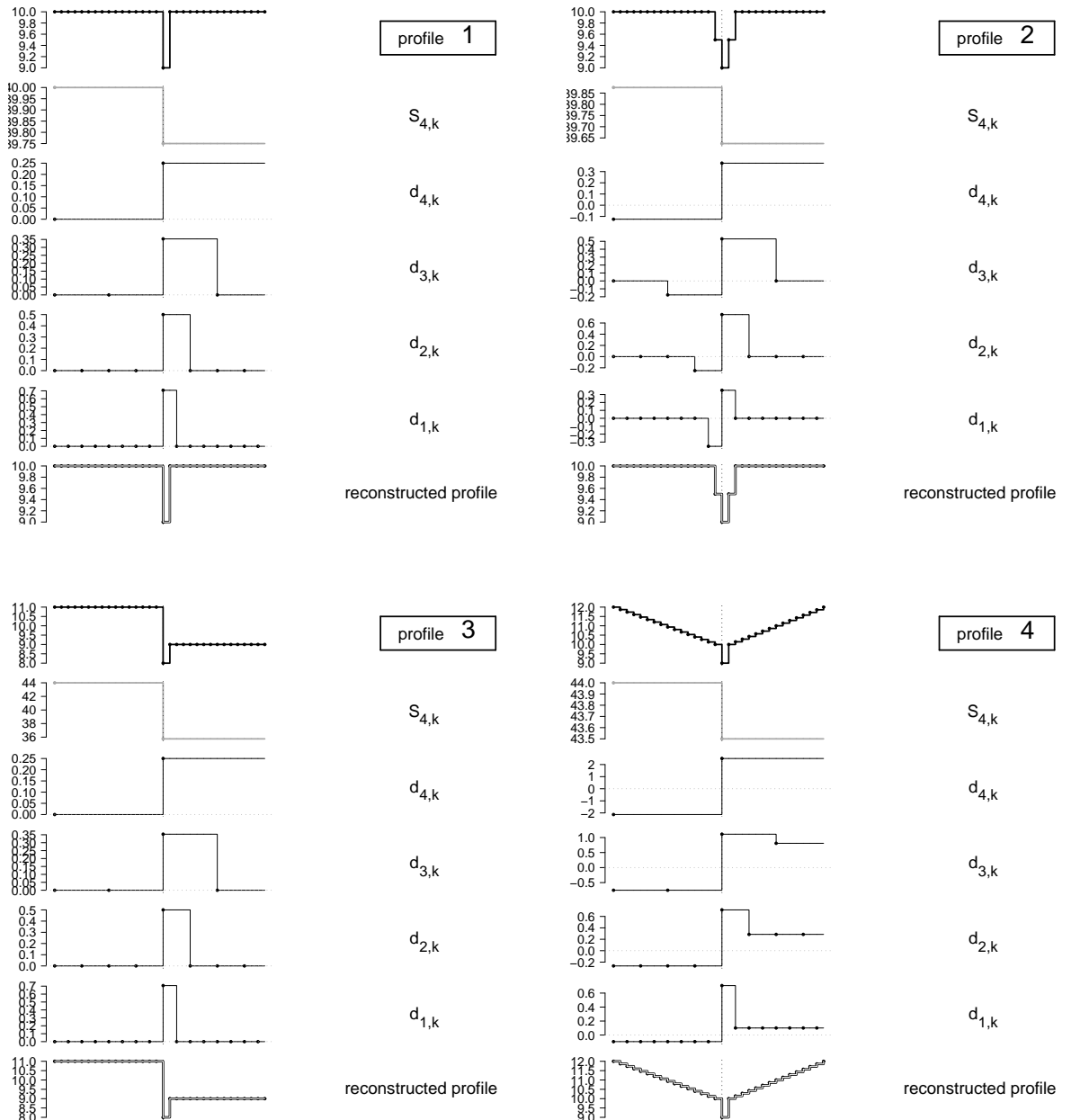


FIG. 4.5 – Comparison of D.W.T coefficients for 4 virtual elevation profiles having a sharp central concavity

- signal trend is captured by coarsest wavelet detailed coefficients  $d_{4,k}$  and the difference between the  $S_{4,k}$  coefficients.
- central concavities alternates a sequence of low and high central detailed coefficients due to Haar  $\phi$  shape. The magnitude of the difference between these coefficients is related to concavity depth.
- central  $d_{1,k}$  sequence is stable on profiles 1 and 3 which are having the same



- central concavity but different trends.
- the magnitude of the difference in the central detailed coefficient sequence for different resolutions is correlated to concavity width as seen when comparing profiles 1 and 2.
  - Despite a known property of D.W.T to exhibit artifacts in the neighborhood of sharp discontinuities, called pseudo-Gibbs oscillations <sup>4</sup> (Coifman et Donoho, 1995), these artifacts do not appear here since the reconstructed profiles are exactly the same as the genuine profiles. This is due to the fact that "when using Haar wavelets, a discontinuity precisely located at location  $\frac{n}{2}$  will lead to essentially no pseudo-Gibbs oscillations" (Coifman et Donoho (1995), p. 126) (for more details see D.W.T and S.W.T comparison in appendix p. 181).

*Dimension reduction : wavelet coefficients selection and P.C.A*

Further, we only retain central wavelet detailed coefficients  $d_{j,k}$  since we are not interested in signal details on borders. We chose to keep information on the profile trend tackled for a part with  $d_{4,1}$ ,  $d_{4,2}$ . We also added a unique variable  $S_{4,1-2}$  equal to the difference between two father wavelets coefficients. Absolute value of  $S_{4,1-2}$  is linearly correlated to profile slope, but its sign depends on  $d_{4,1}$ ,  $d_{4,2}$  sequence. This selection gave us 13 wavelet coefficients.

To facilitate further classifications, a principal component analysis (P.C.A) was performed on the wavelet coefficients. It reduced the dimension of the vectors characterizing the elevation profiles and removed the linear correlations between detailed wavelet coefficients caused by the central concavity (see examples on figure 4.5).

*Watershed algorithm*

Watershed transformations make partitions of a signal into segments with boundaries called crest points for 1D signals. A precise definition of this transform is detailed in Soille (1999). We will only present here the transformation fitting to our problem. Watershed transformation identifies segments by performing successive flooding simulations on local minima in signals. Crests are local maxima. Assumptions placed on ditch concavities by LiDAR profiles are introduced : they have, when they exist, local minima on the 17th value ; Due to coarse profile sampling, there is one and only one local minimum within the concavity.

---

<sup>4</sup>successive undershoot and overshoot of a specific target level due to the lack of translation invariance of the wavelet basis

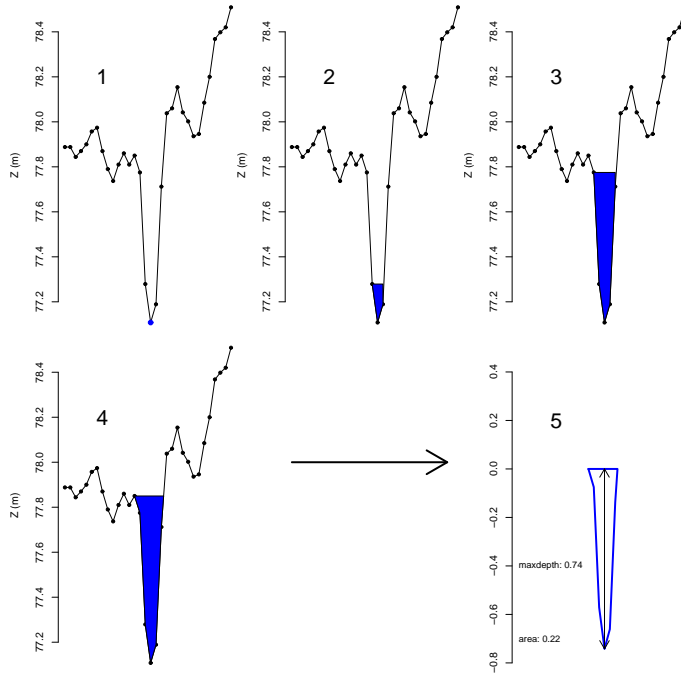


FIG. 4.6 – Example of flooding algorithm 1-flooding starts on the 17th sample (local minima), 2-3-flooding increases while neighborhood elevations increase, 4-flooding ends when crest point is found, 5-concavity is characterized with flooded surface (a translation of -77.85 m is computed on y axis labels in order to show concavity depth in m).

Consequently, we used a flooding algorithm in a simpler way when compared to usual watershed algorithms as explained in figure 4.6 : only the local minimum fixed at the 17 th value of the profile was flooded. Flooding was stopped as soon as a crest point was identified. The maximum water level is then fixed to the altimetry of this crest point and the flooded surface is calculated accordingly.

The calculated flooded surface exhibits a concavity on which we calculate some geometrical variables : area, maximum depth, mean depth, upper width. We selected the area and maximum depth to describe concavity patterns, since ditch concavities should give significant values for these variables.

**Step 3 : classification method** Classifications of LiDAR profiles in  $D, ND$  classes were performed with a segmentation method called CART (Breiman *et al.*, 1984). In brief, the goal of CART is to construct a series of segmentations, i.e. thresholds, which are arranged in a hierarchical manner, namely a tree. The CART algorithm has two stages : Tree growing and tree pruning. The growing stage recursively splits the training samples until the heterogeneity measure is minimized in each terminal node. The rationale behind the heterogeneity measure is that the probability of seeing two different classes in one node should be minimal. Once heterogeneity

measure is minimized for all leaves, the pruning stage begins on an exclusive test set. CART pruning is a risk minimization approach : we keep the subtree that minimizes misclassified individuals of the test set.

*Voir en annexe p. 187 le détail de la méthode CART.*

This popular classification method was selected because 1- it does not need to model a specific distribution of classes in variables space, 2- CART models can be easily interpreted, and 3- CART is a reliable classification method when there are high overlapping classes since pruning stage limits over-fitting (Cortijo et de la Blanca, 1997).

In order to perform CART validation, we used 3 exclusive sample sets which contained the same numbers of  $D$  and  $ND$  samples, successively :

- a calibration or training set. We used Gini index as partitioning criterion for tree node heterogeneity measure.
- a test set to prune the calibrated tree.
- a validation test to assess overall and class-by-class accuracies of pruned CART tree. These accuracies are the only ones mentioned further.

To take account of the variations of results caused by different selections of training sets, CART was performed on a set of learning samples. The use of independent and identically distributed bootstrap samples for tree predictors has been already widely developed in derived CART processes called bagging predictors (Breiman, 1996a) or random forests (Breiman, 1996b). In essence, all of these random subspace methods combined several tree outputs to vote on a final prediction. They were mainly developed to 1- compensate for sampling sensitivity by predictor combining, and 2- identify predictive variable importance for prediction in case of variable correlations. Although we used too a random subspace method for free sampling accuracy calculation, our objective differed, since an easily interpretable and representative tree, corresponding to the mode of all possible trees, was required.

To do so, we repeated the following process : from the 6461 profiles,

1. we randomly sampled  $\frac{1}{3}$  (843) of the  $D$  set and same number of  $ND$  profiles for calibration set,
2. we randomly sampled other exclusive  $\frac{1}{3}$  of  $D$  profiles and same number of  $ND$  profiles for testing set (tree pruning),
3. in the previously non selected profiles, we keep the last  $\frac{1}{3}$  of  $D$  profiles and randomly selected same number of  $ND$  profiles for validation set.

Repeating this process, we built various trees, then computed various confusion matrices and the mean one called the bootstrap matrix. From this set of trees, we kept the tree having the nearest performances to the bootstrap ones, which we called the mean tree.

#### 4.2.1.3 Results and discussion

**4.2.1.3.1 Elevation profile interpretation** Figure 4.7 shows 12 examples of  $D$  LiDAR profiles obtained for real ditches following the procedure described above.

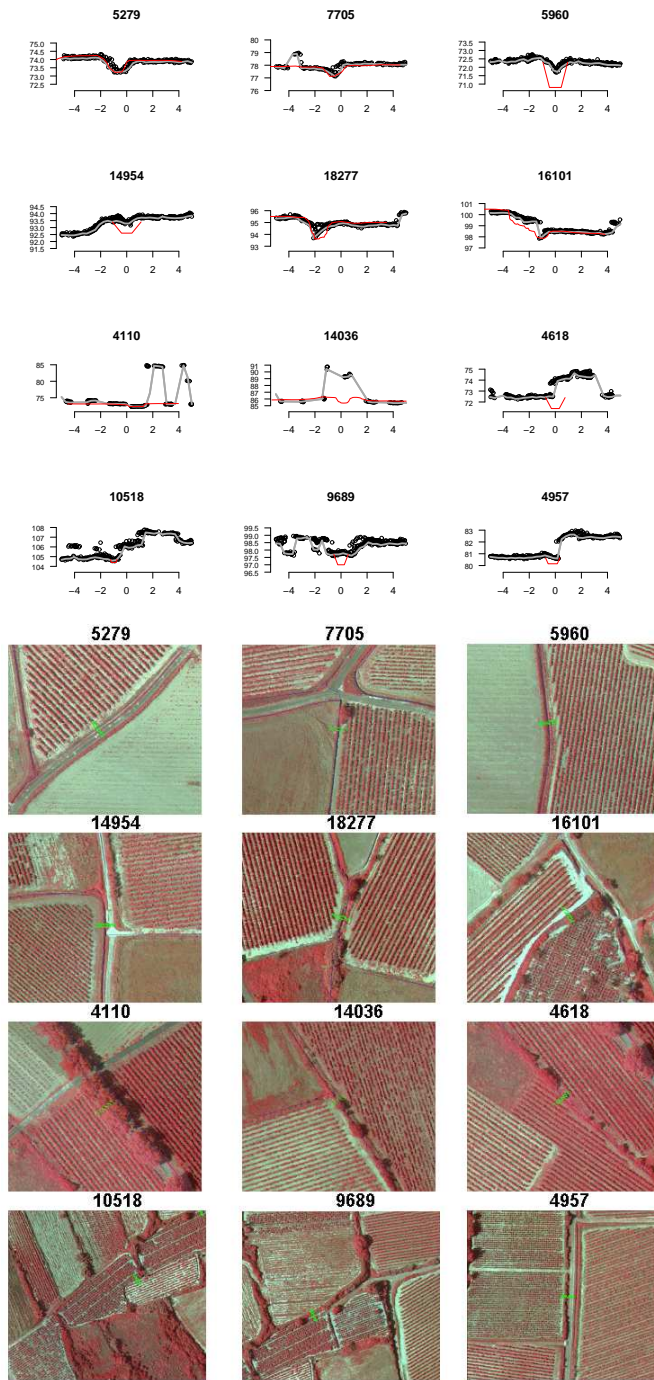


FIG. 4.7 – 12 examples of LiDAR profiles (grey wide curve) estimated from last pulse laser points (black points) compared to actual topographic profiles (red curve). According profiles are located in the center of aerial images (green line).

The first row shows examples of the most favourable situations, i.e. the ditches are located at the boundaries of open fields and in flat areas, where the ditches are clearly marked by a concavity on the LiDAR profiles. However, less favourable situations were encountered over the study area. Examples in the second row show situations where the presence of local relief, here agricultural terraces, predominate over the ditch concavities. In vegetated situations ( third and fourth row) LiDAR last pulses points are not reflected by the terrain, but by vegetation canopies which cause alteration of the resulting elevation profile. For ditches oriented parallel to flight tracks (third row) the spatial sampling of LiDAR points is not optimal, which prevents the detection of any concavity.

Finally, a central concavity was only detected by visual interpretation for 42% of the 2531  $D$  profiles. This prefigures ditch detection rates we can expect with LiDAR profiles.

#### 4.2.1.3.2 Automated profiles classification

**Classification accuracies computation** Calibration, test and validation sets were sampled 1000 times from the 6461 profiles set. We thus calibrated and pruned 1000 CART trees and computed 1000 confusion matrices. From these matrices, we computed mean accuracies (in percents) for  $D$  and/or  $ND$  predictions :  $\bar{t}_{(D,ND)}$  for overall accuracy,  $\bar{t}_D$  for  $D$  accuracy and  $\bar{t}_{ND}$  for  $ND$  accuracy.

**Classification dimension reduction** Classification dimension reduction was only performed when using wavelet coefficients as predictors. We first computed PCA on the 13 wavelet coefficient matrix giving 13 ordered factors called principal components and denoted PC1, PC2, PC3, PC4, ..., PC13. The dimension reduction was performed by selecting the minimum ordered principal component set giving the best significant classification bootstrap overall accuracy  $\bar{t}_{(D,ND)}$ , in an ascending stepwise way : we first computed  $\bar{t}_{(D,ND)}$  when using only PC1 as predictor ; secondly, we computed  $\bar{t}_{(D,ND)}$  using PC1 and PC2 as predictors ; and so on up to  $\bar{t}_{(D,ND)}$  computed with PC1, ..., PC13 as predictors. With this process, we showed that the maximum value for  $\bar{t}_{(D,ND)}$  is obtained when using only the 4 principal components as predictors. These 4 principal components represent 59 % of the 13 wavelet coefficient matrix inertia. Further, only PC1, PC2, PC3, PC4 are kept.

*Voir en annexe p. 184 la signification des composantes principales au travers de la projection, sur les plans des composantes, de profils altimétriques "type" présentés en annexe p. 179.*

Figure 4.8 shows profile locations in the PC planes.  $D$  are depicted in grey and  $ND$  are in black. We added to these planes the median profiles computed from the median wavelet coefficients using equation 4.1. This was done for profiles having extreme PC values, i.e. an absolute PC value greater than one, for each of the eight extreme places in PC planes. For example, on the PC1-PC2 plane on the left, the lower profile down-right is the median profile when clustering profiles having a PC1 greater than 1 and a PC2 smaller than -1 : for this cluster, we computed the median value for each one of the 13 selected wavelet coefficients ; we computed the approximated median profile from equation 4.1 using these 13 median coefficients

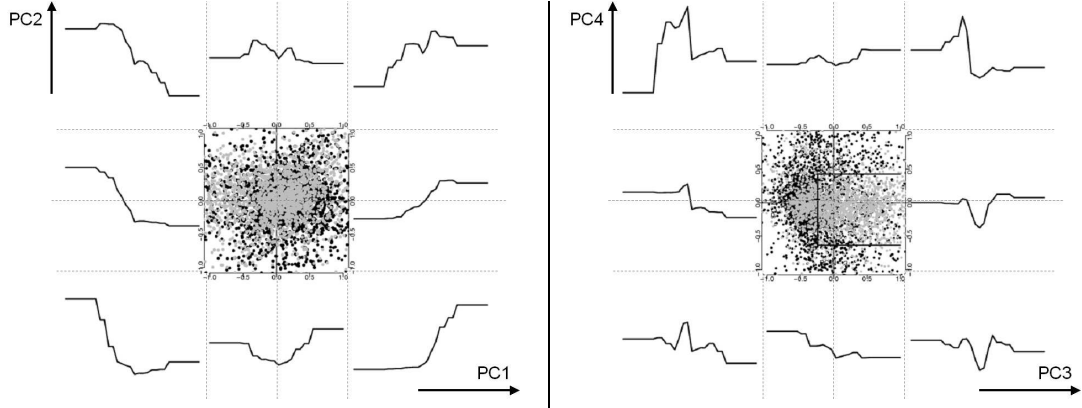


FIG. 4.8 –  $D$  (grey points) and  $ND$  (black points) profiles on P.C.A planes from wavelet coefficients :  $PC1, PC2$  plane on the left and  $PC3, PC4$  plane on the right.

values and zero values for other coefficients. In this figure, it appears clearly, that  $PC1$  and  $PC2$  planes with high contributions of  $d_{4,1}$ ,  $d_{4,2}$  and  $S_{4,1-2}$  are related to a signal trend : slope, a concave or convex trend, and step presence in the middle of the signal.  $D$  and  $ND$  are totally overlapping in this plane. This means there are few specific morphological contexts for ditches. The  $PC3$ - $PC4$  plane is related to central and local concavity with high contributions of the two central wavelet coefficients for the three finest resolutions. It appears, too, that discrimination between  $D$  and  $ND$  is only within the  $PC3$ - $PC4$  plane. In this plane,  $ND$  region looks like a "C" centered on the  $PC3$  axis delineated by the wide black lines on figure 4.8.

**CART Classification accuracies** From the 1000 confusion matrices, we also computed the standard deviation  $\sigma_{t_{(D,ND)}}$  for overall accuracy to give an idea on tree performances variability according to the variable samplings. In addition, we computed the classification results when applying the selected mean tree to the whole 6461 profiles on test zone. These accuracies results are in table 4.1.

		$\bar{t}_{(D,ND)}$	$\sigma_{t_{(D,ND)}}$	$\bar{t}_D$	$\bar{t}_{ND}$
<b>D.W.T</b>	validation sets	67.4	1.50	56.6	78.2
	applied on 6461 profiles	71.3	-	53.7	82.6
<b>Watershed</b>	validation sets	67.0	0.90	45.7	88.3
	applied on 6461 profiles	71.7	-	45.1	88.8

TAB. 4.1 – Bootstrap accuracies on LiDAR profiles classification and classification performances all over study zone using D.W.T or watershed algorithm (in percents)

For the two tested methods of detection of central concavities, the mean ditch omission rate is around 50 % and the mean ditch commission rate is around 15 %. The former rate looks similar to the one obtained above for the visual interpretation of  $D$  profiles. The overall results of D.W.T and Watershed methods do not differ significantly. There is, however, a difference in the balance between commission and

omission rates, D.W.T showing a lower omission rate (-10 %) compensated for by an higher commission rate. Furthermore, D.W.T methods exhibit a higher accuracy variability than watershed methods, which is partly explained by higher dimensions for segmentations with CART using the PCs<sup>5</sup> on the wavelets coefficients.

In the following, only results for the mean trees are used. The mean tree from D.W.T combines additive thresholds on PC3 and PC4 as depicted by the wide black lines in figure 4.8 : *D* profiles are the ones having  $PC3 > -0.24$  and  $-0.64 < PC4 < 0.43$  . As *PC3* is linearly dependent on concavity depth, CART tree threshold on *PC3* is related to a concavity depth greater than 0.4 meters. The mean tree for watershed exhibits a simple threshold : *D* profiles are the ones having concavity surface greater than 0.09 square meters. This can be a surface for a "U" depression of 0.3 by 0.3 meter or for a "V" depression with 0.45 m depth and 0.45 m upper width. For both methods, it appears that only ditches giving significant concavities on their LiDAR profiles deeper than about 0.45 m are detected.

**Predicted classes mapping** Figure 4.9 shows the map obtained by applying the mapping mean tree results on the 6461 profiles of the studied area.

Despite some local differences between maps obtained with D.W.T and watershed methods, omissions and commissions are similarly located. Whatever the method, results are inhomogeneous in space : omissions as well as commissions on ditches are more numerous in the non-flat areas than in the flat areas. These morphological areas are the ones where ground factors that induce misclassifications are the most important, i.e. the presence of vegetation cover and thin ditches located at the bottom of agricultural terraces (see examples on figure 4.7). Commissions are mainly located in forested areas where gaps in the forest canopy may induce profile concavities, too. Finally, commissions and omissions are clearly non-randomly distributed because of the homogeneity of ditch characteristics within a given field boundary.

---

<sup>5</sup>Principal components



FIG. 4.9 – Maps of classification results using wavelets (right) or watershed (left) on study area : blue points are for  $D$  well classified sites, red for ditches omission, orange for ditches commission and black for  $ND$  well classified sites

#### Sensitivity analysis *Accuracy sensitivity to profile azimuth*

As seen in the examples above (see figure 4.7), the variations of flight track azimuth induced variations of LiDAR point sampling which may have consequences on ditches detection. To quantify this effect,  $D$  well-classified rates for each 10 degree range boundary azimuth classes were calculated (figure 4.10).

A 95% confidence band for a null hypothesis corresponding to no azimuth effect was also computed from 1000 simulated classification results obtained by attributing omissions on sites with constant probability equal to overall and isotropic omission rate  $(1 - \bar{t}_D)$  (two black curves on figure 4.10). The black NW-SE wide line is the



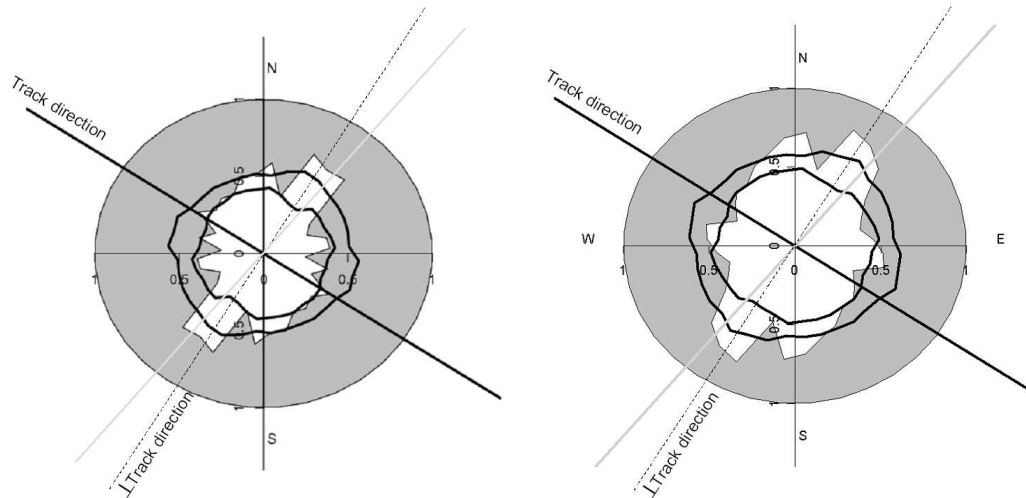


FIG. 4.10 – Plot boundary azimuth effect on  $D$  detection accuracy using wavelets (right) and watershed (left) : the white area is the area of well classified profiles and the grey area is the area of profile misclassifications.

mean flight track direction and the light grey one is the sun azimuth at flight time. Whatever the method, figure 4.10 reveals a higher significant detection rate (0.77 for D.W.T) for ditches perpendicular to flight track azimuth, i.e. profiles parallel to flight track azimuth. Furthermore  $D$  detection rate decreases dramatically when moving away in this situation. However, detection rates for ditches parallel to flight track azimuth, i.e. profiles perpendicular to flight track azimuth, do not really differ from the mean isotropic rate.

#### *Accuracy sensitivity to vegetation cover*

To quantify the role of the vegetation on the detection of ditches that was previously evoked from examples (see figure 4.7), 99  $D$  test sites, evenly distributed over the studied area, were examined. From the field descriptions made during LiDAR surveys, these sites were classified into 4 classes according to their vegetation cover : 1- low vegetation (herbaceous) (LV), 2- high vegetation (shrubs, trees) (HV), 3- both low and high vegetation (LV+HV) and 4- no vegetation (NV).

When counting the mean tree classification results from the 99 sites for each one of the vegetation classes, we obtain the figure 4.11. Whatever the method, we observe high ditch omission rates for sites with vegetation, especially when high vegetation, i.e. shrubs and trees, cover the ditches. When little vegetation exists, detection rates are around 75 %. Thus, it appears clearly that vegetation causes omissions.

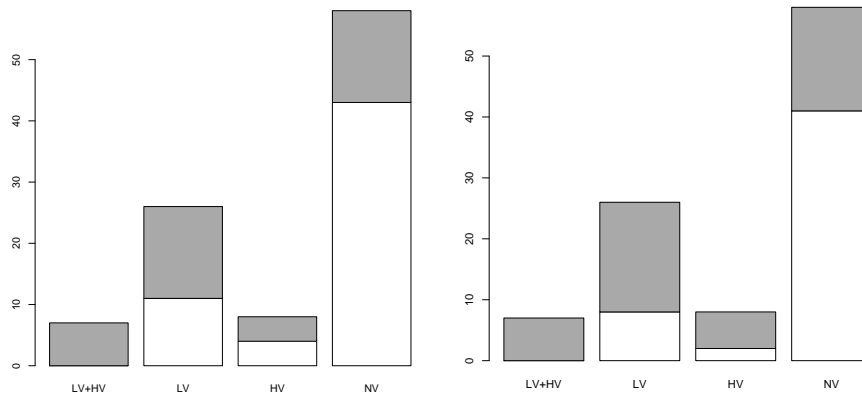


FIG. 4.11 – Vegetation cover effects on ditches detection using wavelets (right) and watershed (left) for 99 sites. For each observed vegetation modality LV, HV, HV+LV, NV, ditches omissions are the grey bars and well detected ditches are the white bars.

#### 4.2.1.4 Discussion

In this study, we designed a methodology for automated ditch network detection from LiDAR data in Mediterranean vineyard landscapes. This methodology is based on LiDAR elevation profile classification using curve shape analysis methods. We assessed the accuracies of the proposed methodology using two different curve shape analysis methods : wavelet transform and a specific watershed algorithm.

For both methods, we obtained a satisfactory overall profile classification accuracy of about 70 % with unbalanced confusions : ditch omissions are numerous (around 50 %) and ditch commissions are few (around 15 %). The results for ditch omissions are similar to the ones obtained when performing visual classifications of profiles. Moreover, as the performances of the two methods were not significantly different, we presume that the low performances for ditches detection are caused for a major part by the LiDAR data. But to better compare methods performances or to better asses the parts in the results coming from methods and from LiDAR data, additional work is required. For example, this work could be based on topographic profiles classification estimated from simulated LiDAR data with different LiDAR spatial sampling design, different laser pulse energies, different LiDAR footprint resolutions. Of course, for this goal, the implementation and the validation of models that simulate LiDAR echoes distribution with different LiDAR system configurations or different surface environments is first required.

Regarding results sensitivity, two factors governing ditch omissions must be stressed : 1- trees, shrubs and herbaceous vegetation covering ditches at survey time masked the terrain with the LiDAR resolution and power we used and 2- the LiDAR sampling rate across flight tracks in the scanning system was not dense enough. Therefore, substantial progresses in ditch detection can be expected by the use of a LiDAR survey at leaves-off vegetation periods or with smaller footprints and with a scanning system that gives a more regular and isotropic spatial sampling design. Despite the rather low ditch detection performances, the poor commission rate looks

promising. Further research is required to fill the gaps in detection, by adding external knowledge on ditch network architectures.

The transposability of the proposed methodology to other agrarian linear features depends on two conditions : 1-the features must be located on field boundaries and 2-they must correspond to an elevation discontinuity that can be identified in LiDAR profiles. For the former, LeCoeur *et al.* (2002) reminds us that most interesting structural agrarian linear features in western Europe bocage landscapes are on field boundaries. Furthermore, the use of geographical digital cadastral databases is no longer a limitation, since these databases are becoming available (<http://www.cnig.serveur-1.net/fiches/27cadas.htm>) (Trias-Sanz et Deseilligny, 2004; Durupt et Taillandier, 2006). For the latter, wavelet transform or other specific morphological algorithms can be use and compared since they can both fit to various shape detections or characterizations.

## 4.2.2 Classification binaire des arêtes du lattice cadastral à partir des profils LiDAR classés

### 4.2.2.1 Introduction

Dans cette sous-section, on explore puis propose une méthode de classification des arêtes du lattice parcellaire en mode binaire (modalités fossés  $D$  et non-fossé  $ND$ ). Pour chaque arête, cette classification s'effectue à partir des profils LiDAR classés tels que présentés en page 110 et situés sur cette arête. Elle aboutit à une homogénéisation des classes  $D$  et  $ND$  par arête. On parlera ensuite de "lissage" pour cette homogénéisation par arête. Cette échelle d'homogénéisation des résultats de télédétection se justifie car un fossé superpose généralement en totalité une arête du parcellaire cadastral, très rarement une partie seulement. Seuls les résultats des profils classés suivant leur transformée en ondelettes seront considérés ici.

La littérature sur des méthodes d'homogénéisation de résultats de classifications imprécis, incertains, pour un niveau de segmentation donné et avec des segments variables en dimensions est assez réduite : en télédétection, ces méthodes portent généralement sur des règles de majorité des pixels classés par segment : classification de l'occupation du sol en mode parcellaire (Raclot et Colin, 2005), agrégation des états de surface à l'échelle d'une parcelle agricole (Waasenaar, 2001). Ces règles sont spatiales, dans le sens où elles ne tiennent pas compte de la répartition spatiale initiale des classes à l'intérieur des segments qui, dans certains cas (et par hypothèse dans le nôtre), peut être informatif.

La méthodologie proposée pour la recherche d'une règle optimale de lissage se déroule en deux étapes : 1—pour chaque arête du lattice parcellaire, la répartition spatiale des résultats de classification est synthétisée suivant des indices spatiaux simples traduisant leur "texture" ; 2—un modèle statistique de classification (CART) basé sur ces indices est utilisé pour construire cette règle. CART est choisi cette fois pour sa capacité à utiliser des variables prédictives qualitatives aussi bien que quantitatives, et toujours, pour sa facilité d'interprétation. Cette méthodologie aboutit à un arbre de classification utilisant les indices construits qui traduit la règle de lissage. La méthode mise en oeuvre est présentée sur les résultats de classification des profils suivant le modèle "moyen" basé sur les coefficients d'ondelettes des profils (voir p. 110). L'application aux résultats de profils classés suivant d'autres modèles, plus "extrêmes", sont présentés en annexe page 197. Ces résultats sont comparés par rapport à un lissage classique qui s'appuie sur une simple règle d'agrégation majoritaire.

### 4.2.2.2 Données

Le jeu de données utilisé pour calibrer la règle de lissage correspond à celui situé sur la même zone test **B** de  $2 \text{ km}^2$  que celui utilisé dans la sous-section précédente (voir p. 96). On dispose sur cette zone **B** de 6461 sites (profils) classés ( $ND = 0$  ou  $D = 1$ ) suivant le modèle moyen sélectionné à partir des coefficients d'ondelettes (voir page 110) et distribués sur 869 arêtes. Ces profils classés sont ceux présentés en haut de la figure 4.14. Une première impression visuelle montre des résultats ponctuels hétérogènes sur les arêtes du lattice, notamment par rapport à la pente du terrain : ils semblent plus agrégés par arête dans la dépression centrale de la zone et paraissent plus erratiques sur les zones pentues.

Le modèle CART est calibré sur la zone **B** puis validé sur une zone test de  $1 \text{ km}^2$ , dénommée zone **C**, située à l'est de la zone **B** (voir situations p. 49). Sur cette

zone **C**, on dispose de 2999 sites classés sur 569 arêtes du lattice parcellaire (voir en bas à gauche de la figure 4.14). Ces arêtes sont inégalement réparties entre une faible proportion située sur une zone pentue et boisée au sud-est et, a contrario, une très forte proportion située en zone plane de dépression. Sur la zone **C**, le taux de linéaire d'arêtes du lattice en fossé est élevé.

### 4.2.2.3 Méthodes

**4.2.2.3.1 Choix des variables descriptives par arête** Pour tenir compte de la répartition spatiale des résultats, les indices de texture suivants sont introduits : *Suite1max* qui comptabilise par segment le nombre maximum de séquences consécutives (1,1) et *Suite0max*, l'équivalent pour les séquences (0,0) (Fig. 4.12).

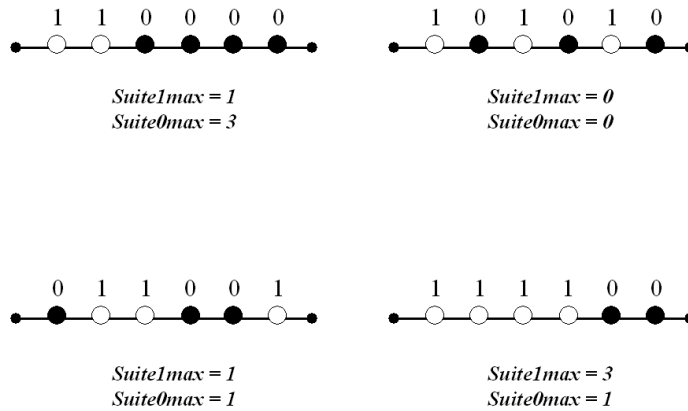


FIG. 4.12 – Illustrations des indices *Suite1max* et *Suite0max* par arête.

On pourrait penser qu'il faille normaliser les indices *Suite0max*, *Suite1max*, i.e. les diviser par le nombre de profils classés par arête, représentatif de la longueur de l'arête. Les différentes causes d'omission de la présence de fossés sur les profils altimétriques LiDAR identifiées précédemment (couverture végétale, orientation du fossé...) n'ont pas de raison de générer un nombre d'omissions proportionnel à la longueur de l'arête. Il en est de même pour les erreurs de sur-détection de fossés (commission) sur les profils altimétriques LiDAR, également liées à la présence de végétation sur l'arête. De ce fait, nous garderons ces indices par arête non normalisés par la longueur de l'arête.

**4.2.2.3.2 Arbre de classification** La méthode de classification utilisée pour traduire la règle de lissage par arête du lattice, CART, est la même que celle utilisée dans la sous-section précédente. En revanche, les échantillons statistiques sont cette

fois-ci beaucoup plus réduits (rapport de 5 à 1 environ), pour un même nombre de classes recherchées (2). Pour cette raison, nous n'emploierons pas de méthode de sélection de modèle suivant une procédure de ré-échantillonnage aléatoire. Plus classiquement, le modèle sera construit suivant 2 phases élaborées sur deux échantillons différents : un de calibration et un de test. En revanche, le critère maximisé pour ce lissage par classification étant un taux de linéaire d'arêtes bien classées, chaque individu statistique (arête) est affecté d'un poids proportionnel à sa longueur. La sélection de l'arbre de classification est réalisée classiquement en deux phases :

1. une phase de "calibration" d'un arbre profond (nombre de feuilles important) sur l'échantillon de la zone **B** (869 individus). Cet arbre profond est construit suivant un processus de segmentation qui maximise le gain d'homogénéité des feuilles calculé sur l'indice d'hétérogénéité de Gini, en tenant compte du poids de chaque échantillon, et avec comme critère d'arrêt de la segmentation, un nombre minimum de 10 individus par feuille (voir l'explication de l'algorithme CART en annexe p.187).
2. si l'arbre calibré contient plus de 2 feuilles, une phase de "test" sur l'échantillon de la zone **C** (569 individus) est réalisée. En partant de l'arbre construit en phase 1, l'arbre est élagué progressivement en diminuant une à une le nombre de feuilles terminales jusqu'à n'obtenir qu'une partition binaire (2 feuilles). A chaque étape d'élagage, les matrices de confusion et les taux de linéaire total d'arêtes bien classées sur la zone **C** (et **B**) sont calculés. L'arbre élagué qui maximise ce taux est sélectionné in fine.

Cet arbre sélectionné représente la règle de lissage retenue.

#### 4.2.2.4 Résultats

L'arbre issu de la phase de calibration est un arbre à 19 feuilles. Les évolutions par élagages successifs des taux de linéaires bien classés (global,  $[D, ND]$ ,  $D$ ,  $ND$ ) sur les échantillons des zones **B** et **C** sont représentées sur la gauche de la figure 4.13 : les taux globaux  $[D, ND]$  sont représentés par les lignes larges, les lignes simples sont celles représentant les taux  $ND$  et les lignes en pointillé sont celles représentant  $D$ . Les taux calculés sur la zone **B** sont en noir et ceux calculés sur la zone **C** sont en rouge. L'optimum est atteint pour un arbre à 4 feuilles (ligne grise verticale). Cet arbre est celui présenté à droite de la la figure 4.13. Sur chaque noeud, on peut visualiser le critère de seuillage retenu et pour chaque feuille terminale, la classe dominante : 0 ( $ND$ ) et 1 ( $D$ ). Selon ce modèle, une arête est classée  $D$  si :

$$(\text{Suite1max} \geq 2) \text{ ou } (\text{Suite1max} \geq 1 \text{ et } \text{Suite0max} = 0) \quad (4.2)$$

et  $ND$  sinon.

Le tableau 4.2 récapitule les taux de linéaires d'arêtes bien classées pour les zones **B** (calibration) et **C** (test) suivant l'équation 4.2 par comparaison aux taux de profils LiDAR bien classés. Ces taux sont directement comparables puisque les profils

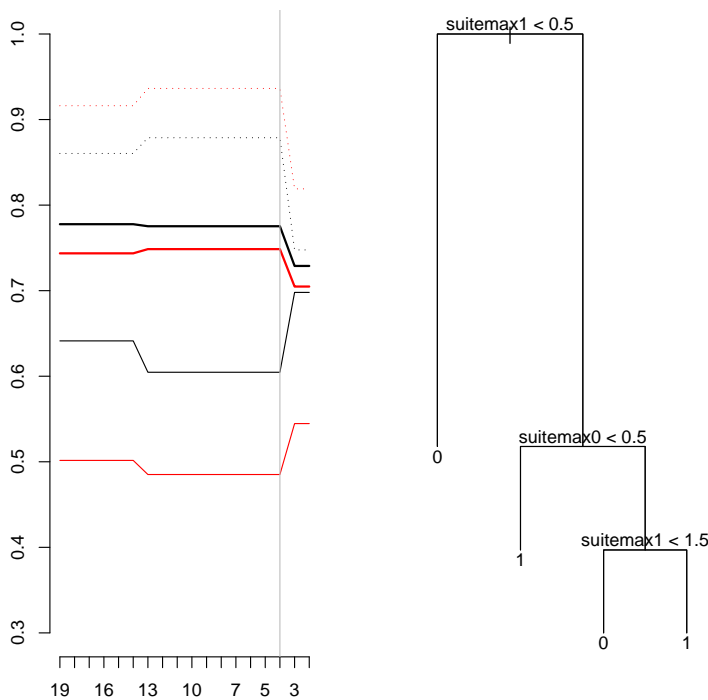


FIG. 4.13 – Evaluation du taux de linéaire d’arêtes bien classées suivant la complexité (nombre de feuilles) de l’arbre (Gauche) ; Arbre à 4 feuilles sélectionné (Droite).

LiDAR sont disposés pseudo-régulièrement sur les segments du lattice parcellaire. Leur nombre est donc représentatif d’une longueur. En dernière ligne, sont calculés les linéaires d’arêtes bien classées si une simple règle majoritaire est appliquée (l’arête est classée  $D$  si le nombre de profils  $D$  qui la composent est supérieur au nombre de profils  $ND$  et inversement).

Sur ce tableau, on peut voir que le lissage suivant des indices de texture (Eq. 4.2) améliore significativement les taux globaux de linéaire d’arêtes bien classées sur les deux zones. Cependant, ces taux sont très comparables à ceux obtenus suivant une agrégation par simple règle majoritaire. Les différences portent sur l’équilibre entre les classes pour les deux classes (différence dans les résultats du linéaire d’arêtes  $D$  et  $ND$ ). Les résultats du lissage suivant l’équation 4.2 sur les zones 1 et 2 sont présentés à droite de la figure 4.14 (Droite : profils LiDAR classés ; Gauche : résultats lissés par arête suivant les mêmes codes couleurs : noir pour  $ND$  bien classé, bleu pour  $D$  bien classé, orange pour  $ND$  classé en  $D$  et rouge pour  $D$  classé en  $ND$ ).

	$(D, ND)$	$D$	$ND$
zone B avant lissage	<b>71</b>	54	83
zone C avant lissage	<b>72</b>	56	86
<b>Lissage par Eq.4.2</b>			
zone B après lissage	<b>78</b>	60	88
zone C après lissage	<b>75</b>	49	94
<b>Lissage par règle majoritaire</b>			
zone B après lissage	<b>77</b>	57	90
zone C après lissage	<b>74</b>	56	88

TAB. 4.2 – Pourcentages du linéaire des arêtes bien classées avant et après lissages sur les zones B (calibration) et C (test).

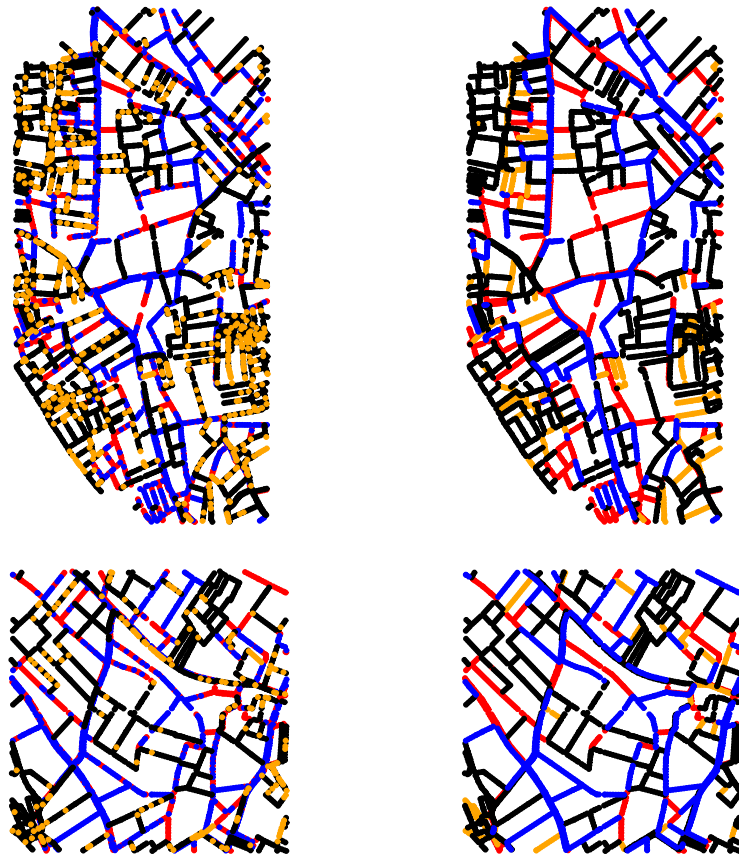


FIG. 4.14 – Comparaison des erreurs de classification avant et après lissage (Eq. 4.2) sur la zone B (Haut) et C (Bas) (même signification des codes couleurs).



#### 4.2.2.5 Conclusion

Dans cette sous-section, nous avons étudié l'homogénéisation à l'échelle des arêtes de lattice parcellaire des profils LiDAR classés ponctuellement sur ces arêtes. Cette homogénéisation a pour objectif de classer ces arêtes en mode binaire fossés ( $D$ ) et non-fossé ( $ND$ ). L'homogénéisation qui correspond à un lissage spatial des profils LiDAR classés  $D$  et  $ND$  a été réalisée par arête suivant un arbre de classification élaboré à partir d'indices de texture spatiale des profils LiDAR classés  $D$  et  $ND$ .

Bien que, par manque d'échantillonnage, nous n'ayons pas réellement réalisé de validation externe de la règle de lissage présentée (utilisation des zones **B** et **C** pour élaborer cette règle), on peut conclure sur des résultats de lissage suivant des indices de texture spatiale similaires à ceux issus d'une simple règle d'agrégation spatiale par classe majoritaire. Ce résultat montre une faible structure spatiale stable des classes de profils LiDAR interne aux arêtes du lattice parcellaire. Quoiqu'il en soit, cette homogénéisation par arêtes permet globalement d'améliorer le linéaire d'arêtes classées suivant un gain moyen d'environ 3 à 4% du linéaire total. Ce gain n'est pas équivalent pour les classes  $D$  et  $ND$  : il est plus marqué pour les arêtes  $ND$ . Le taux de détection des arêtes "fossé" ( $D$ ) reste assez faible (50 à 60 %).

Une étude de sensibilité de la règle de lissage optimale basée sur les indices de texture *Suite0max* et *Suite1max* a été réalisée pour d'autres modèles de classification préalable des profils LiDAR (voir l'annexe page 197). Toutes les règles présentent le même critère principal de segmentation au niveau de la racine de l'arbre CART : au moins une séquence de profils consécutifs ( $D, D$ ) sur une arête tend à classer l'arête comme arête  $D$ . Cette simple règle semble robuste et généralisable à divers résultats de classification des profils LiDAR, et ce, suivant des performances qui semblent plus stables qu'avec une règle plus classique de classification des arêtes par classe majoritaire. Cependant, pour juger plus avant de cette propriété, il faudrait tester les performances de ces différentes règles d'homogénéisation par validation externe, en les appliquant sur d'autres sites échantillonnés, dans des contextes différents.

L'ensemble des arêtes classées en mode binaire est une étape qui permet ensuite de reconstituer un réseau arborescent au sein du lattice cadastral, reconstitution conditionnée (passant) par les arêtes classées "fossé" ( $D$ ) du lattice.

## 4.3 Algorithme de simulation conditionnelle d'un réseau de fossés dans le lattice des limites cadastrales

### 4.3.1 Introduction

L'objectif de cette section est de présenter une méthode de reconstitution d'une arborescence dans un lattice orienté lorsque l'exutoire ainsi qu'un ensemble d'arêtes disjointes de l'arborescence sont fixées. Appliquée au lattice des limites de parcelles cadastrales, cette méthode simule des réseaux hydrographiques artificiels (fossés, buses et éléments de voiries connectés), conditionnées par une connaissance partielle et incertaine du réseau : la rivière principale, connue, et les résultats de télédétection des fossés, incertains, de la section précédente. Pour cette connaissance partielle du réseau, on parlera plus tard d'arêtes "données".

D'un point de vue méthodologique, plusieurs méthodes de simulation d'arborescences sont proposées dans la littérature. Ces méthodes de simulation sont pour la plupart non conditionnelles (sans sous-ensemble de l'arborescence fixé) et se réalisent dans un espace peu contraint (Manna, 1998; Broder, 1989). Quelques méthodes conditionnelles récentes ont été proposées mais restent difficilement transposables dans notre cas. Là encore, c'est la particularité du support des simulations, i.e. un lattice orienté, qui contraint fortement l'espace des possibles des simulations et limite cette transposabilité. Par conséquent, une méthode spécifique de simulation conditionnelle d'arborescences est proposée. Cette méthode s'appuie sur un algorithme qui réalise des cheminements aléatoires (random walk) amont-aval dans le lattice suivant des critères topographiques des noeuds du lattice. Cet algorithme se déroule en plusieurs étapes et contraint l'arborescence simulée à présenter une distribution spatiale de ses sources spécifique, spécificité à modéliser en fonction du type de milieu étudié.

Les arborescences sont donc simulées aléatoirement mais sous différentes contraintes (conditionnement) : 1- l'ensemble des arêtes données imposent les points de départ et d'arrivée des cheminements aléatoires dans le lattice vers l'aval, vers l'exutoire ; 2- la structure du lattice (topologie), notamment l'orientation du lattice par la topographie, limite les possibilités de cheminement aléatoire amont-aval. Une autre caractéristique du lattice contraint les cheminements vers l'aval : le lattice de départ des limites de parcelles cadastrales est un ensemble de sous-lattices déconnectés, séparés par des routes et chemins ; les possibilités de cheminement aléatoire entre îlots de parcelles cadastrales sont donc liées aux possibilités d'interconnexion entre ces îlots parcellaires ; 3- une distribution spatiale des sources spécifique impose aux arborescences simulées une configuration topologique et spatiale.

Cette section est organisée dans l'ordre des points suivants : 1- la bibliographie sur les indices descriptifs de l'architecture d'une arborescence (topologie et géométrie), indices à la base de la plupart des méthodes de modélisation et de simulations d'arborescences, est présentée. Ensuite, les quelques méthodes de simulation conditionnelle de réseaux relevées dans la littérature sont évoquées ; 2- nous présentons schématiquement l'algorithme proposé et les différentes étapes et paramètres qui le définissent ; 3- le cas d'étude, les différentes contraintes des simulations, i.e. le modèle de distribution de sources retenu, la structure du lattice et de la topographie sont présentés ; 4- la distribution d'un ensemble de simulations par comparaison au "vrai" réseau est étudiée suivant des indices et critères d'architecture d'arborescences pré-

sentés au point 1 ; 5– nous concluons sur le réalisme des simulations réalisées suivant ce simple processus aléatoire sous contraintes puis sur les limites et la transposabilité de la méthode proposée.

## 4.3.2 Bibliographie : Simulations d'arborescences

### 4.3.2.1 Indices descriptifs de l'architecture d'arborescences

Ne seront pas présentés ici les indices associés aux nombreuses lois physiographiques qui relient le réseau hydrographique aux surfaces drainées (Lois de Schumm, Hacks, Langbein (Dodds et Rothman, 1999)). En effet, la surface drainée en chaque point du réseau est difficilement accessible à des échelles fines et sur un milieu aménagé (Comoretto, 2003).

En revanche, on s'intéressera à l'architecture des arborescences selon deux caractéristiques : la topologie, i.e. la manière dont se branchent aux confluences les arêtes du réseau et la distribution des sources qui en découle puis la géométrie, i.e. les dimensions métriques ou angulaires des arêtes ou branchements.

#### 4.3.2.1.1 Topologie d'une arborescence

**Taxonomie de Strahler** Ce sont essentiellement les hydrologues qui ont créé tout un ensemble d'indices relatifs à la topologie d'une arborescence : Horton, Strahler, Shreeve. Nous allons ici présenter la méthode d'Horton-Strahler, méthode de classification largement utilisée, même en dehors de l'hydrologie : étude de l'architecture d'arborescences en botanique, en neuro-physiologie, en tectonique du sous-sol, en informatique et en biologie moléculaire. . . (Janey, 1992, Chapitre 2).

##### *Classification de Horton-Strahler*

La taxonomie de Horton-Strahler (Strahler, 1952), abrégée par la suite HS, subdivise un réseau hydrographique en un ensemble d'arêtes ordonnées hiérarchiquement. On peut voir cette méthode comme une méthode d'élagage successifs d'un arbre, applicable quelque soit le degré supérieur des confluences (Fig. 4.15). On définit tout d'abord une arête source, d'ordre  $w = 1$ , comme une arête liant une source à la première confluence rencontrée en aval. Ensuite, on élague le réseau de toutes ces arêtes "sources", puis on identifie sur ce réseau élagué les nouvelles arêtes sources, qui deviennent les arêtes d'ordre  $w = 2$ , et ainsi de suite jusqu'à l'ordre maximal atteint lorsqu'une arête a pour noeud terminal l'exutoire du bassin considéré. Cet ordre maximal  $w_{max}$  est l'ordre du bassin.

La figure 4.15 représente schématiquement cette hiérarchisation sur un arbre binaire (haut) puis sur un arbre présentant des confluences ternaires<sup>6</sup> (bas), non rares sur des réseaux artificiels.

---

<sup>6</sup>de degré supérieur égal à 3

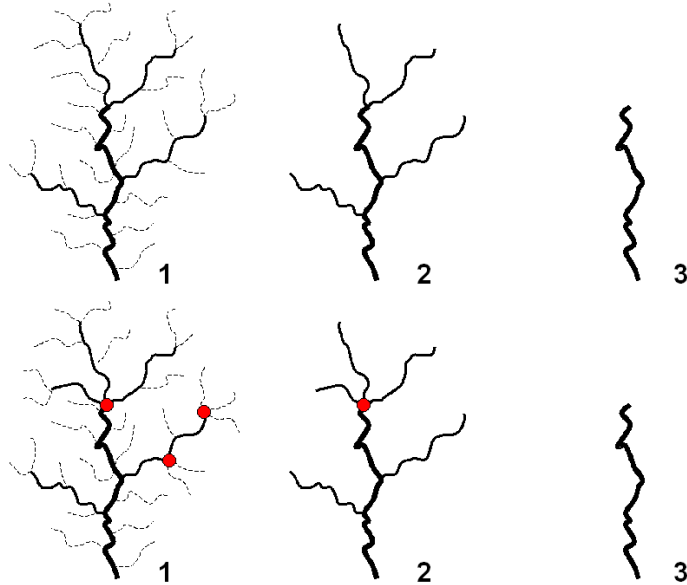


FIG. 4.15 – Taxonomie de Horton-Strahler pour une arborescence binaire (Haut) et exemple pour des confluences ternaires indiquées en rouge (Bas).

**Distribution des sources** Un autre critère étudié pour décrire la topologie d’une arborescence en hydrologie (Peckham, 1995) est la distribution (nombre) des sources du réseau (Istanbulluoglu *et al.*, 2002; Moussa, submitted). Cette distribution liée à celle des arêtes d’ordre  $w = 1$  quantifie les arêtes terminales des arborescences. Dans le cas des réseaux artificiels des paysages cultivés, cette distribution ne peut-être simplement regardée dans l’espace géographique. Elle doit être regardée dans l’espace du lattice parcellaire puisque les sources sont un sous-ensemble des noeuds du lattice. Un noeud du lattice peut quelquefois correspondre à plusieurs sources<sup>7</sup>.

**Matrice de ramification** La distribution des sources du réseau permet de quantifier pour partie la grandeur d’une arborescence mais elle ne permet pas de décrire le processus de branchement aux confluences de cette arborescence. Des arborescences aux allures très différentes peuvent présenter le même nombre de sources. Pour ce faire, on peut utiliser comme seconde grandeur topologique, une matrice de ramification (Janey, 1992, chapitre3). Cette matrice descriptive des branchements d’une arborescence est utilisée pour caractériser nombre de structures arborescentes naturelles : les éclairs électriques, les digitations visqueuses aux interfaces de liquides de différentes viscosités . . .(Janey, 1992). Les principes et calculs d’une matrice de ramification sont détaillés ci-après.

Dans le cas d’un arbre binaire, on calcule pour chaque confluence, terminaison initiale d’une arête d’ordre  $w$  son bi-ordre, c’est à dire les deux ordres HS des arêtes ayant cette confluence comme terminaison finale. Ce bi-ordre est soit de type  $(w - 1, w - 1)$ , soit de type  $(w, i)$  avec  $i < w$ . Sur la figure 4.16, on représente sur la droite

<sup>7</sup>D’un noeud du lattice cadastral, peuvent démarrer plusieurs fossés selon des directions opposées. Dans la réalité, ces sources sont souvent séparées de quelques mètres.

un arbre dont les confluences sont en rouge. L'épaisseur du trait de chaque arête est relatif à son ordre HS  $w$  : les traits pointillés représentent les arêtes  $w = 1$ , les traits plein, les arêtes  $w = 2$  et les traits épais, les arêtes  $w = 3$ . Pour chaque confluence, les bi-ordres sont comptabilisés dans la matrice en haut à droite. Trois exemples de bi-ordre sont situés dans cette matrice, respectivement de haut en bas : un bi-ordre (1, 1) donnant une arête d'ordre 2, un bi-ordre (2, 1) donnant une arête d'ordre 2 et un bi-ordre (2, 3) donnant une arête d'ordre 3. Les branchements aux 24 confluences sont ainsi comptabilisés sur cette matrice.

Du dénombrement précédent, on note  $p_{w,i}$ , la probabilité pour une confluence d'ordre  $w$  d'avoir un bi-ordre  $(w, i)$  ou  $(w - 1, w - 1)$ . La matrice de ramification d'un arbre  $G$  est la matrice stochastique de dimension  $\max(w) * (\max(w) - 1)$  dont les composantes sont les  $p_{w,i}$ . La somme de chaque composante en ligne est égale à 1. Si l'exemple de la figure 4.16 est repris, on obtient le dénombrement des bi-ordres présentés dans la matrice en haut à droite, transformée en matrice de ramification (en bas à droite) en subdivisant chaque case par la somme en ligne (12 pour chaque ligne).

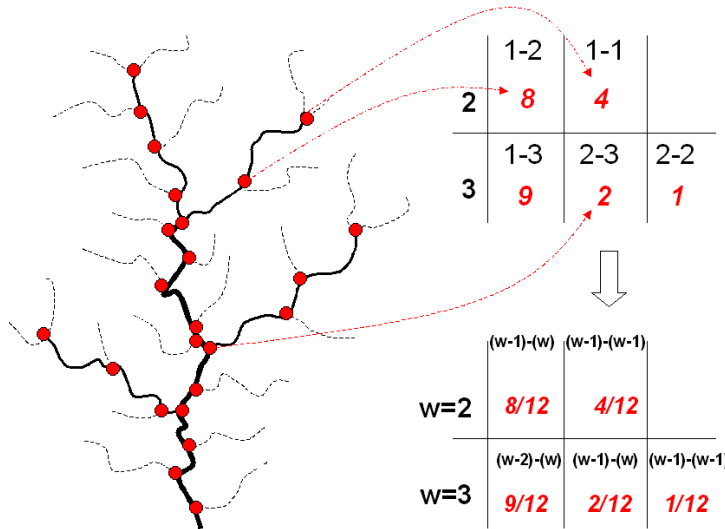


FIG. 4.16 – Exemple de calcul d'une matrice de ramification

Ces matrices, stochastiques, peuvent être utilisées pour simuler des topologies d'arborescences suivant une procédure de simulation allant de l'aval vers l'amont, de l'exutoire vers les sources.

**4.3.2.1.2 Géométrie d'une arborescence** Plusieurs critères sur la géométrie d'une arborescence peuvent être utilisés :

- la distribution des angles au niveau des confluences, utilisés notamment en synthèse d’images d’arbres (Janey, 1992, chapitre3). Dans notre cas, les angles sont non aléatoires et fixés par la structure du lattice.
- une loi de longueur des arêtes<sup>8</sup> de l’arborescence entre confluences, généralement en fonction de  $w$  (Mendes-France, 1981; Dodds et Rothman, 1999) (voir différents modèles en annexe p.201). Ces lois expriment souvent des propriétés d’invariance d’échelle du réseau du réseau, i.e. la longueur moyenne des arêtes d’ordre HS  $w$  est une fonction puissance de  $w$ .

#### 4.3.2.2 Lois et simulations non-conditionnelles d’arborescences

**4.3.2.2.1 Approches fractales** Des lois fractales associées à des procédures aléatoires, sont largement utilisées pour simuler d’une manière non-conditionnelle des arborescences : par exemple pour simuler des réseaux vasculaires (Dokoumetzidis et Macheras, 2003) ou pour simuler des systèmes racinaires (Salas *et al.*, 2004).

*De multiples lois basées sur quelques hypothèses d’invariance d’échelle*

En hydrologie, la classification de Horton-Strahler est souvent utilisée pour inférer la structure de réseaux hydrographiques naturels, notamment suivant des lois dites de similitudes ou d’invariance d’échelle, i.e. liées au comportement fractal des réseaux hydrographiques naturels et des bassins associés.

Par exemple, les lois de Horton (Dodds et Rothman, 1999) étudient des rapports adimensionnels qui traduisent, lorsqu’ils sont stables, une dimension fractale ou un spectre de dimensions fractales (multi-fractals) : rapport constant des nombres d’arêtes d’ordre  $w$  et  $w + 1$  ou rapport constant des longueurs moyennes des arêtes d’ordre  $w$  et  $w + 1$  . . . , et ce quelque soit la valeur de  $w$ .

Un autre exemple vient des lois de Tokunaga basées sur les statistiques des confluences de bi-ordres de type  $(w, i)$  avec  $i < w$  (voir en annexe p. 201).

Dodds et Rothman (1999, p. 12) ont montré que ces lois multiples sont en fait des expressions différentes des trois mêmes hypothèses d’invariance d’échelle spatiale, à savoir :

1. une densité de drainage uniforme dans l’espace, c’est à dire, pour tout bassin versant, un rapport constant entre le linéaire de réseau et la surface du bassin ;
2. une auto-affinité du réseau, ce qui traduit une sorte de sinuosité uniforme du réseau (Fig. 4.17) : pour chaque bassin, la longueur  $l$  de chaque rivière principale du bassin peut s’écrire, pour tout ordre  $w_{max}$  du bassin comme  $l = L^d$ , avec  $L$  "longueur" du bassin (Fig. 4.17) et  $d$  paramètre fractal<sup>9</sup> (Molz *et al.*, 2004).
3. une auto-similarité de la structure du réseau, c’est à dire qu’une partie du réseau est identique au réseau complet. Cette auto-similarité se traduit classiquement par l’invariance des rapports de Horton égaux à un paramètre fractal  $D$ .

---

<sup>8</sup>Une arête a ici des terminaisons qui sont soit des sources, des confluences ou l’exutoire.

<sup>9</sup>généralement compris entre 1 et 2

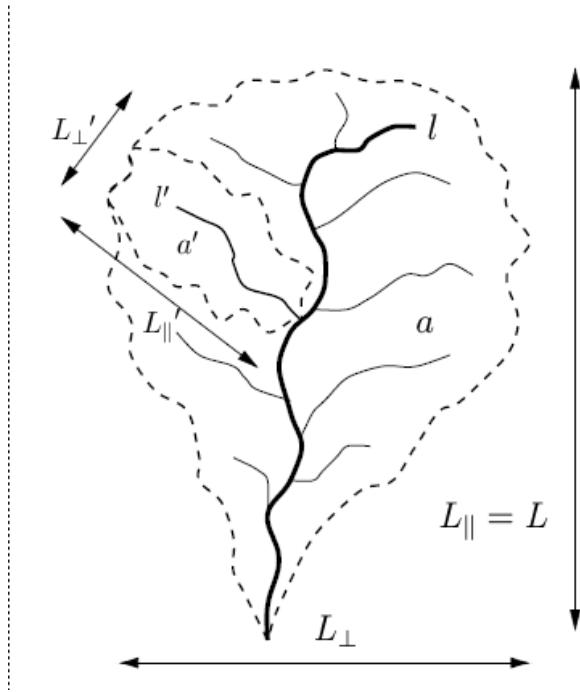


FIG. 4.17 – Exemple de deux bassins emboîtés de surfaces  $a$  et  $a'$ , avec deux rivières principales de longueurs  $l$  et  $l'$  et deux longueurs équivalentes  $L$  et  $L'$ , extrait de Dodds et Rothman (1999).

#### *Des réseaux aléatoires comme situation topologique de référence*

La plupart des lois précédemment évoquées ont été testées et inférées sur de "vrais" réseaux hydrographiques pour de grands bassins, relativement plats (Mississippi par exemple). Takayasu (1990) et Molz *et al.* (2004) ont tous deux montré que des arborescences (réseaux hydrographiques) aléatoires reproduisent les mêmes lois d'invariance d'échelles que celles observées pour les vrais réseaux. Ces arborescences aléatoires sont, par exemple, celles de Scheidegger (Scheidegger, 1991), arborescences binaires construites en suivant la ligne de plus grande pente dans un paysage aléatoire<sup>10</sup>. Les statistiques sur ces arborescences purement aléatoires fournissent donc une situation de référence, une hypothèse nulle, pour tester la topologie d'une arborescence (Dodds et Rothman, 1999). Cette idée générale peut-être reprise dans le cadre des réseaux artificiels.

En revanche, lorsque l'on s'intéresse aux réseaux artificiels sur de petits bassins ruraux, réseaux en amont des grands systèmes hydrographiques, les hypothèses à la base des lois d'invariance d'échelle sont pour certaines difficiles à assumer, notamment une auto-similarité du réseau et une densité de drainage uniforme. Cet écart aux hypothèses d'invariance d'échelle est d'autant plus fort que, pour ces réseaux, la morphologie et la topographie du terrain 'organisent' fortement le réseau. On

<sup>10</sup>Paysage aléatoire construit suivant une altitude aléatoire entre 0 et 1 affectée à chaque noeud d'un lattice triangulaire régulier; altitude ensuite remaniée pour ne plus présenter de minima locaux

s'éloigne par conséquent des réseaux purement aléatoires de Scheidegger, construits sur terrain plan, bruité et sans structure (Niemann *et al.*, 2001). Nous ne retiendrons donc pas une approche fractale comme approche de modélisation et de simulation de l'architecture du réseau. Nous chercherons à employer des critères plus généraux.

**4.3.2.2 Approches basées sur les matrices de ramification** On peut voir sur la figure 4.18 des exemples d'arbres binaires simulés en 3D, extraits de Janey (1992). Ces arbres sont simulés à partir des matrices de ramification, représentées au dessous de chaque arbre. L'arbre parfait (gauche) est un arbre que ne contient que des bi-ordres ( $w - 1, w - 1$ ) : il présente aussi des propriétés fractales. Un arbre aléatoire de  $n$  confluences (milieu) est un arbre tiré au hasard parmi tous les arbres binaires à  $n$  confluences possibles ( $\frac{2n!}{n*(n+1)!}$  combinaisons). Un arbre aléatoire croissant (droite) est une suite croissante d'arbres aléatoires à une confluence dont la racine est choisie aléatoirement parmi les sources de l'itération précédente.

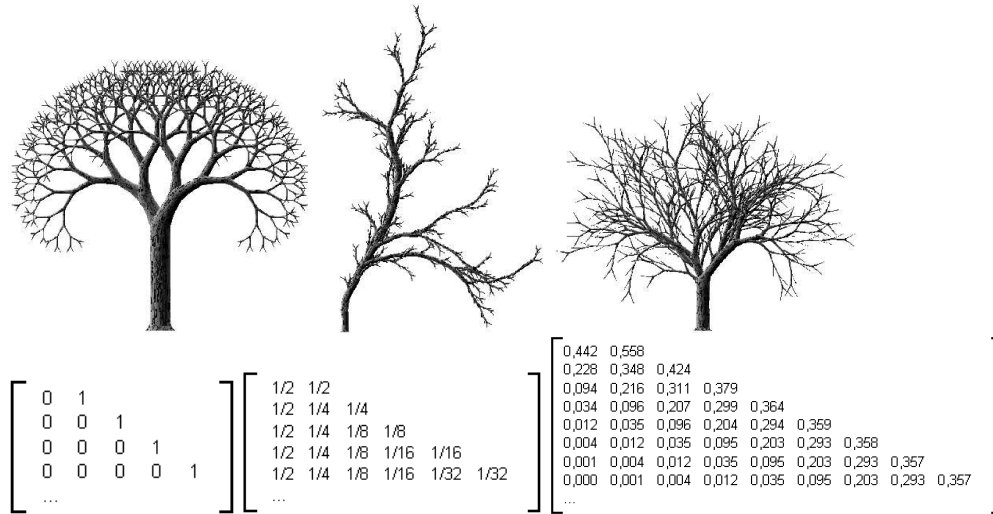


FIG. 4.18 – Exemples de matrices de ramification et d'arbres binaires correspondants (en 3D), extrait de Janey (1992) : arbre parfait, arbre aléatoire, arbre aléatoire croissant.

### 4.3.2.3 Simulations conditionnelles de réseaux

Toutes les méthodes présentées auparavant sont utilisées pour réaliser des simulations stochastiques non-conditionnelles de réseaux arborescents. Quelques références de méthodes de simulation conditionnelle stochastique de réseaux existent dans la littérature. Généralement, ces méthodes s'appuient sur un modèle de géométrie stochastique pour reconnecter des portions de réseaux, lesquelles portions de réseaux sont extraites de télédétection. On peut notamment citer le modèle développé initialement par Stoica (2001). Ce modèle correspond à un processus de polygones marqué, la marque portant entre autre les azimuts des arêtes de réseau. La méthode mise en oeuvre initialement pour la reconnection de routes (réseau maillé)



partiellement détectées sous couvert forestier, optimise "l'alignement" de segments du réseau.

Très récemment, Lacoste (2004) a développé une approche similaire sur des réseaux hydrographiques arborescents naturels de plaine et de zone humide. Cette approche part d'un tronc (portions larges de réseau aval) identifié sur des images et s'appuie dans un second temps sur un processus de polylignes hiérarchiques pour ramifier ce tronc vers l'amont. Le processus de polylignes simulé (algorithme MCMC à saut réversible) respecte 1- une structure arborescente et 2- des critères géométriques notamment de distance euclidienne minimum entre rivières du réseau (pour homogénéiser la densité de réseau et des sources dans l'espace). On interdit dans ce cas des rivières parallèles très proches les unes des autres. Même si l'objectif semble proche du notre, ces méthodes semblent difficilement exploitables dans notre cas : 1- les azimuts des arêtes à reconnecter sont déjà fortement conditionnés par la structure du lattice; 2- on raisonne dans un espace qui est celui du lattice cadastral, les distances euclidiennes dans l'espace géographique ont peu de sens; 3- dans le cas des réseaux artificiels, deux rivières peuvent être très proches (de part et d'autre d'une route) et donc un éloignement minimum ne peut être un critère géométrique pertinent et enfin 2- la reconnection est également liée à des considérations topographiques et topologiques (orientation du lattice et du réseau), autres que géométriques.

#### 4.3.2.4 Bilan

En résumé, les méthodes de simulation non conditionnelle d'arborescences extraites de la littérature sont basées soit 1- sur une hypothèse d'invariance d'échelle du réseau (dimension fractale), hypothèse que nous ne pouvons vérifier dans notre cas de réseaux artificiels (pas assez de données) mais qui semble difficilement admissible; soit 2- sur des matrices de ramifications qui représente la structure de branchement associée et des lois géométriques de longueurs de segments (Janey, 1992). L'introduction d'une matrice de ramification type pourrait être envisagée dans une procédure de simulation conditionnelle aval-amont d'une arborescence dans le lattice cadastral (tirage aléatoire des bi-ordres suivant les probabilités par ligne de la matrice de ramification). Cependant, nous ferons l'hypothèse que les degrés de liberté sont trop réduits sur un lattice donné (système déjà fortement contraint) pour pouvoir converger vers une solution qui respecte cette matrice et les autres contraintes pour les simulations.

En revanche, de cette revue bibliographique, nous pouvons retenir plusieurs critères et grandeurs qui permettent de comparer l'architecture d'une arborescence, sa topologie et sa géométrie, à savoir :

- les matrices de ramification pour représenter la hiérarchie des branchements. On peut notamment tester le caractère aléatoire des branchements d'une arborescence par comparaison de sa matrice de ramification à celle d'un arbre aléatoire.
- des critères géométriques simples : une densité de drainage globale (somme du linéaire total du réseau ramené à la surface totale du bassin), voire une densité de drainage "locale" par sous-secteur du lattice. Dans notre cas, la densité calculée doit être le rapport entre le linéaire de l'arborescence et le linéaire du lattice pour ce sous-secteur.

Les méthodes de simulation non-conditionnelle existantes, issues de la géométrie stochastique, sont quant-à-elles, trop spécifiques pour pouvoir être simplement

adaptées à notre problématique. Une autre méthode doit donc être proposée.

### 4.3.3 Méthode de simulation conditionnelle d'arbres orientés dans un lattice orienté

#### 4.3.3.1 Schéma général

L'algorithme proposé est initialisé à partir d'un ensemble de données, modèles et paramètres :

- un lattice (des îlots cadastraux séparés par des routes) orienté par un M.N.T (Fig. 4.19-a),
- une partie d'arêtes données et l'emplacement de l'exutoire (Fig. 4.19-b) ; les arêtes données provenant de télédétection sont le résultat de la section précédente, i.e. une détection ponctuelle et incertaine du réseau dans l'espace du lattice qui est ensuite "filtrée" par arête du lattice,
- un modèle "ad-hoc" de distribution spatiale des sources du réseau (Fig. 4.19-c).
- de quelques paramètres supplémentaires qui seront introduits au fur et à mesure de la présentation des étapes.

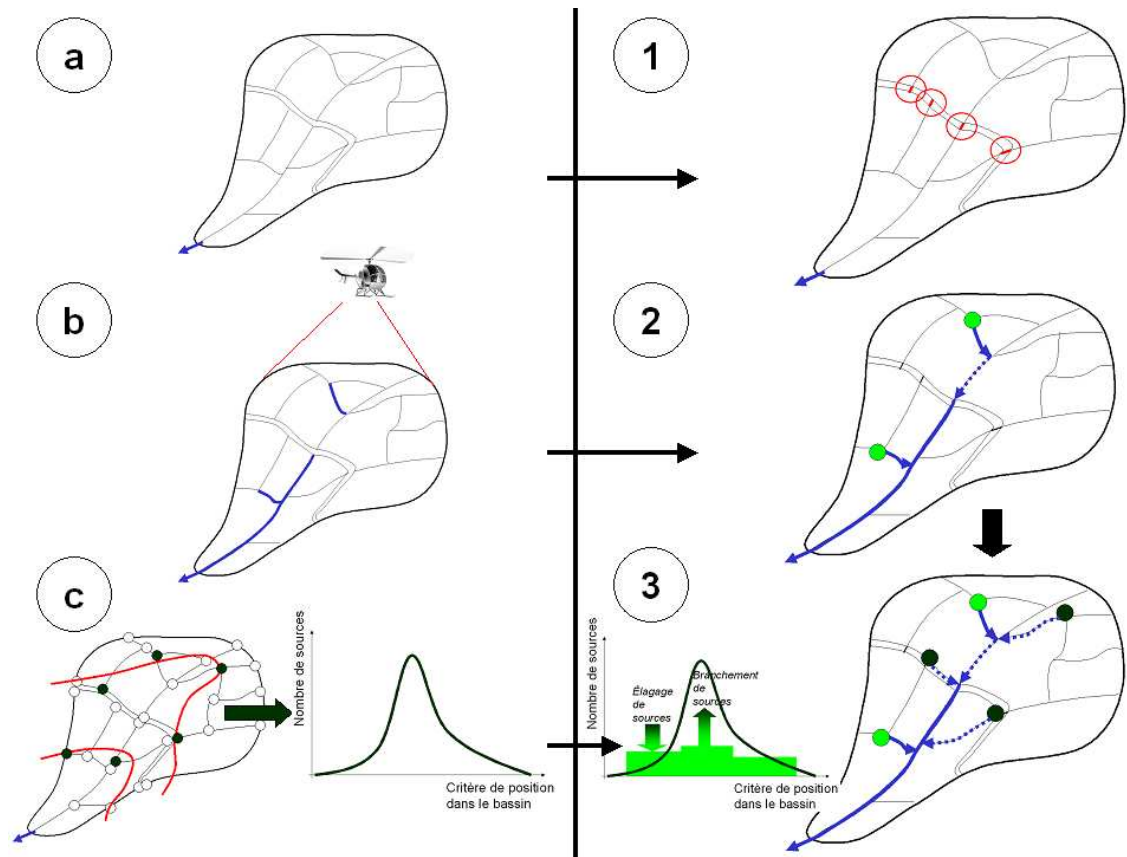


FIG. 4.19 – Données, modèles (gauche) et étapes (droite) de l'algorithme

Cet algorithme se déroule en trois étapes (figure 4.19-droite) :

1. une étape de "densification" du lattice, pour interconnecter les îlots parcellaires (Fig. 4.19-1),
2. une étape de construction d'un "tronc" de l'arborescence par reconnection des arêtes données (Fig. 4.19-2),
3. à partir de la réalisation d'un modèle de distribution spatiale des sources, on reconnecte des sources supplémentaires au tronc précédent, ou au contraire, on élague le tronc d'arêtes sources (Fig. 4.19-3).

Ces étapes sont détaillées ci-après.

In fine, l'algorithme simule aléatoirement une arborescence triplement conditionnée :

- par les arêtes données du réseau : on part et on passe par ces arêtes ;
- par la structure et l'orientation du lattice qui imposent les degrés de liberté du caractère aléatoire des cheminements amont-aval ;
- par une distribution spatiale des sources sur le lattice (bassin) étudié.

Dans l'algorithme, il n'y a pas de conditionnements topologiques de l'arborescence autres que ceux venant de la distribution des sources et de la structure du lattice orienté. De plus, il n'y a aucune contrainte imposée sur la géométrie de l'arborescence (longueurs des segments, angles. . .).

Pour modéliser la distribution spatiale des sources, vu la particularité de l'espace considéré (lattice cadastral), on s'intéresse notamment à la distribution de la proportion  $P = S/N$ , densité de sources dans l'espace du lattice, avec  $S$ , cardinal de l'ensemble des sources du réseau considéré ou à simuler et  $N$ , cardinal des noeuds du lattice dans lequel simuler l'arborescence (fixé). En effet, une première cause de variabilité apparente de la densité des sources dans l'espace géographique vient de la densité du lattice (nombre de noeuds), i.e. de la taille du parcellaire. Suivant le milieu étudié, cette distribution peut être modélisée dans l'espace suivant une "co-variable" qui peut être un index de position  $c$  dans l'espace étudié ou une autre variable connue partout, par exemple :

- l'altitude du terrain. Dans ce cas, les sources sont organisées par tranches d'altitude qui correspondent à des affleurements hydromorphes entre couches sédimentaires ;
- l'orientation du terrain (hydromorphies différentielles en fonction d'effets de versants et de l'orientation de la pente : adret, ubac. . .) ;
- une co-variable géologique ou climatique, à plus petite échelle (plus grands domaines).

#### **4.3.3.2 Étape 1 : Densifier le lattice par reconnection des îlots parcellaires**

Afin de pouvoir assurer une continuité du réseau entre les parties amont et aval du bassin, et compte tenu de la présence d'îlots cadastraux "isolés" (par des éléments de voirie), on crée pour toute paire de noeuds du lattice initial dont la distance à "vol d'oiseau" dans l'espace géographique est inférieure à un paramètre  $\Delta h$ , une arête (si elle n'existe pas) (Fig. 4.20). Cette étape permet de connecter tous les îlots parcellaires voisins et de trouver un cheminement des eaux de n'importe quel point du lattice vers l'exutoire.

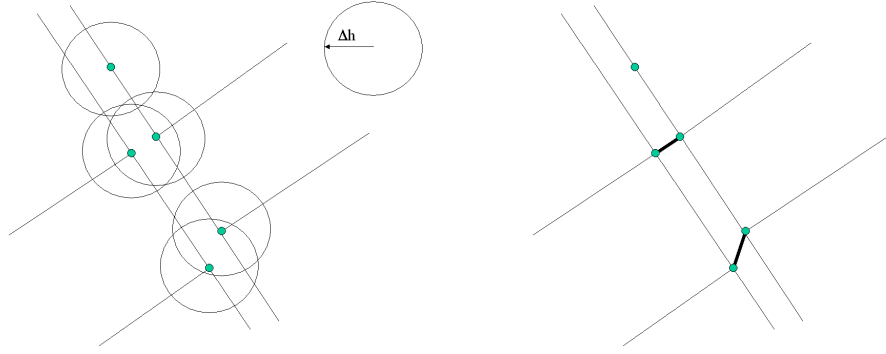


FIG. 4.20 – Ajout d'arêtes sur le lattice entre noeuds (points verts) des îlots parcelaires suivant un critère de distance  $\Delta h$

#### 4.3.3.3 Étape 2 : Former un "tronc" orienté à partir des arêtes données

Pour cette étape, on introduit un paramètre de dénivelée dans le lattice, noté  $\Delta Z$ , et appelé ensuite paramètre de "désorientation du lattice". Ce paramètre joue un double rôle : 1– il tient compte du bruit sur l'altitude des noeuds du lattice cadastral, compte-tenu notamment de l'imprécision du M.N.T ; 2– il correspond également à une dénivelée négative tolérée pour une orientation amont-aval égale au sens d'écoulement sur une arête du lattice, ce qui limite l'importance des critères d'altitude pour l'orientation amont-aval (au sens écoulement) du lattice en zone plane. Autrement dit, l'orientation d'une arête du lattice entre deux noeuds  $N_1, N_2$  d'altitudes respectives  $Z(N_1), Z(N_2)$ , devient univoque  $\overrightarrow{(N_1, N_2)}$ , si et seulement si :  $Z(N_1) > Z(N_2) + \Delta Z$ .

L'étape 2 se décompose en phases successives :

1. dans un premier temps, on regroupe les arêtes du lattice données, contiguës entre elles, formant ainsi des branches déconnectées les unes des autres. Pour chacune de ces branches, on sélectionne un exutoire : si  $Z_{min}$  désigne l'altitude la plus faible des noeuds de cette branche (noeuds extrémités des arêtes composant la branche), on sélectionne aléatoirement un noeud parmi ceux dont l'altitude est inférieure à  $Z_{min} + \Delta Z$ . On classe ensuite chacune de ces branches suivant un ordre croissant de l'altitude de leurs exutoires (Fig. 4.21).
2. A partir de chaque exutoire de branche, et suivant l'ordre d'altitude croissante des exutoires précédents, on effectue un cheminement aléatoire par altitude décroissante (à  $\Delta Z$  près) dans le lattice dépourvu des arêtes connues jusqu'à atteindre un noeud d'une autre branche ou l'exutoire du bassin. Lors de ce cheminement aléatoire, on construit au fur et à mesure un ordre amont-aval des noeuds traversés qui est défini par l'orientation du cheminement. A chaque pas de ce cheminement, on se trouve sur un noeud  $N_f$  d'altitude  $Z_f$  ; on sélectionne alors tous les noeuds voisins<sup>11</sup>  $N_{t_i}$  d'altitude  $Z_{t_i}$ , noeuds potentiels à l'aval de  $N_f$  sur le lattice tel que  $\forall i = 1, \dots, n, Z_f + \Delta Z \geq Z_{t_i}$ . Lorsque le nombre de

<sup>11</sup>Noeuds appartenant à une arête où  $N_f$  est l'autre extrémité et exclusifs des noeuds déjà parcourus lors du cheminement

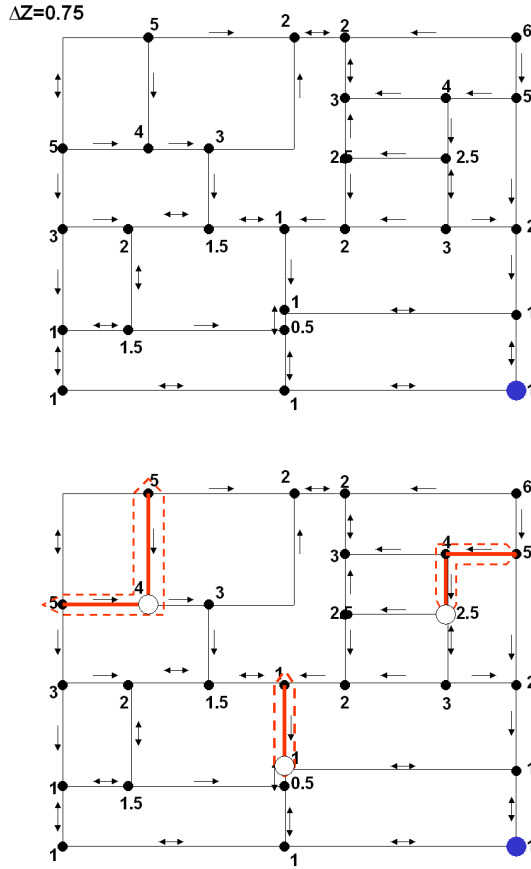


FIG. 4.21 – Haut : Exemple de lattice et orientations résultantes (flèches noires) suivant un critère de "désorientation"  $\Delta Z = 0.75$  (altitude notée à chaque noeud). Bas : Identification des branches d'arêtes données et de leur exutoire (points blancs). L'exutoire  $r_G$  donné est le point bleu

noeuds aval potentiels est supérieur ou égal à 2, on tire alors aléatoirement un de ces noeuds suivant quatre possibilités (Fig. 4.22) :

- une équiprobabilité. La probabilité de sélectionner  $Nt_i$  est  $\frac{1}{n}, \forall i \in 1, \dots, n$
- une probabilité proportionnelle aux dénivelées : on privilégie un écoulement suivant la plus grande pente. Ces différences de probabilité peuvent être accentuées suivant un paramètre multiplicatif des différences de dénivelées. Ce cas n'est envisageable qu'avec des données fiables et précises sur les dénivelées.
- une probabilité proportionnelle à l'alignement vers l'exutoire : on privilégie un tropisme vers l'exutoire. Comme précédemment, ces différences de probabilité peuvent être accentuées suivant un paramètre multiplicatif des différences d'angles. Ce cas n'est envisageable que pour un bassin de forme concentrique.
- une probabilité proportionnelle à la distance minimale dans le lattice entre

le point amont et une autre branche connue selon un chemin qui passe par le point aval considéré. Dans ce cas on favorise les plus courts chemins et on minimise globalement le linéaire de réseau.

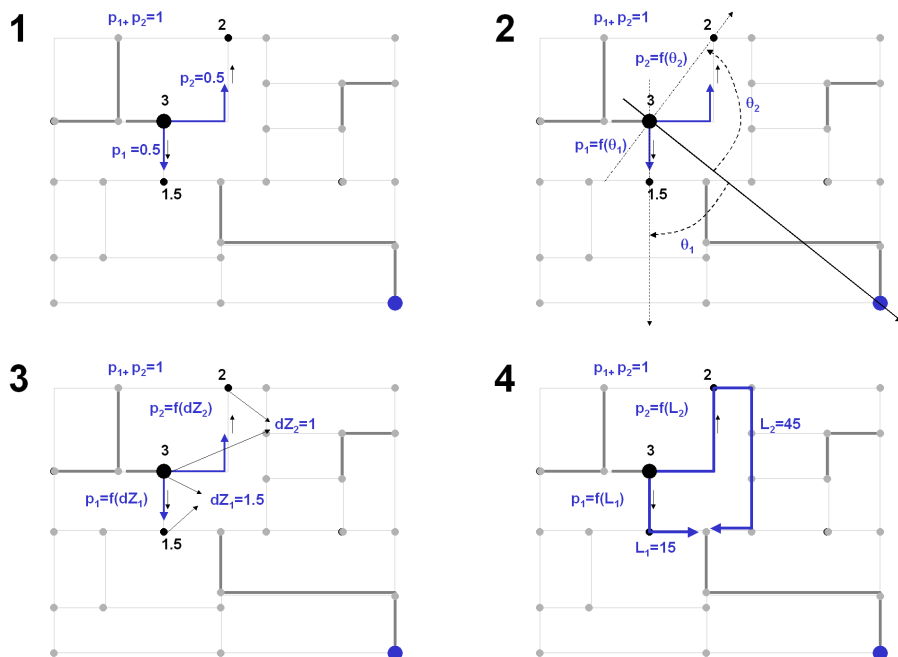


FIG. 4.22 – Différentes règles de tirage d'un noeud aval : 1– probabilité uniforme, 2– probabilité favorisant un cheminement dans la direction de l'exutoire, 3– probabilité favorisant la plus grande pente, 4– probabilité favorisant le plus court chemin.

Si, lors de ce cheminement aléatoire, on finit par une voie sans issue sur un puits (Fig. 4.23 en bas), alors on repart du dernier noeud traversé où un tirage aléatoire a été effectué, vers une autre voie et ainsi de suite. ... Afin de limiter les temps de calculs de la simulation, on limite le nombre de ces itérations (retours vers l'amont) à un maximum, noté  $R_{max}$ . De ce fait, certains chemins amont-aval peuvent correspondre in fine à des voies sans issue avec pour terminaison un puits<sup>12</sup> correspondant à une noeud de degré inférieur nul.

3. On identifie les voies sans issues (éliminées par la suite), et le réseau amont-aval connecté et convergent vers l'exutoire du bassin  $r_G$ , réseau appelé "tronc" (Fig. 4.23).

<sup>12</sup>Un paramètre de forçage des puits  $\Delta P$  est également introduit afin de "réhausser" ces puits d'une valeur éventuellement différente de  $\Delta Z$

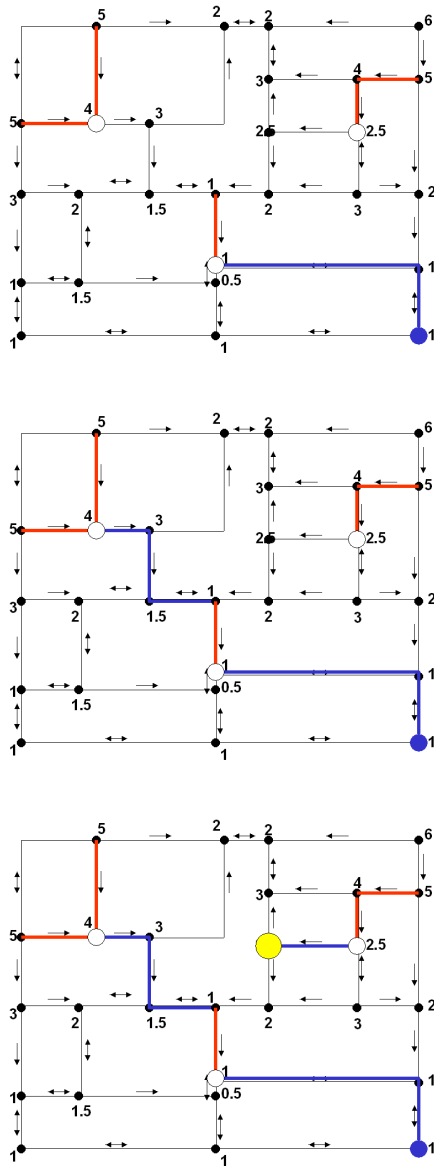


FIG. 4.23 – Exemple de cheminement amont-aval aléatoire depuis 3 exutoires de branches données (en bleu et dans l'ordre) : seul le troisième aboutit à une voie sans issue (puits) représentée en jaune.

#### 4.3.3.4 Étape 3 : Élagage et branchements sur le tronç pour converger vers une distribution de sources

Sur le tronç obtenu précédemment (qui ne contient pas forcément toutes les arêtes données du fait de la présence de puits), on identifie les sources et les confluences.

On calcule ensuite la distribution des sources du tronç suivant l'index spatial

retenu (tranche d'altitude, orientation du terrain. . .), noté  $c(s)$  : on obtient ainsi pour des intervalles  $c_1, c_2, \dots, c_n$  de l'index  $c(s)$ , des nombres de sources notés  $N1(c_i)$ . Sur la base du modèle de distribution des sources suivant  $c(s)$ , on réalise une simulation de cette distribution. De cette simulation, on calcule pour chaque même intervalle  $c_i$  de l'index  $c(s)$ , un nombre de sources  $N2(c_i)$ . On cherche ensuite à modifier le tronç obtenu précédemment de manière à ce que la distribution de ses sources finisse par équilibrer la réalisation du modèle précédent :  $\forall c_i, N1(c_i) = N2(c_i)$ .

Dans un premier temps, lorsque  $N1(c_i) > N2(c_i)$ , on élague le tronç par suppression de "branches sources" sélectionnées aléatoirement. Chaque branche source sélectionnée est alors une séquence contiguë d'arêtes d'ordre HS  $w = 1$  partant d'une source dont la position  $c(s)$  est comprise dans l'intervalle considéré  $c_i$ . Lorsque  $N2(c) > N1(c)$ , au contraire, on complète le réseau en ajoutant des sources sélectionnées aléatoirement parmi les noeuds du lattice dont la position  $c(s)$  est comprise dans l'intervalle  $c_i$  considéré. Depuis ces nouvelles sources, on reproduit le cheminement amont-aval aléatoire dans le lattice présenté dans l'étape 2 : on crée ainsi de nouvelles branches qui se connectent au tronç. A chaque étape, on recalcule les distributions de sources  $N1(c_i)$  qui se modifient au fur et à mesure. On continue cette phase de "branchements" et "élagages" jusqu'à égalité des deux distributions de sources :  $\forall c_i, N1(c_i) = N2(c_i)$ .

A la fin de cet algorithme itératif "d'élagage-branchement", on obtient un réseau arborescent qui limite l'élagage des arêtes données et qui converge vers un modèle de distribution spatiale de sources.

**4.3.3.4.1 Outils** L'algorithme proposé a été développé sous **R** à partir de la table sur S.I.G ARCINFO<sup>13</sup> des arcs et des noeuds du lattice contenant en information attributaire 1- pour les noeuds : l'altitude, les coordonnées géographiques ; 2- pour les arcs : un identifiant binaire marquant leur appartenance à l'ensemble des arêtes 'données'. La simulation d'un centaine d'arborescences sur un bassin de  $3 \text{ km}^2$  prend environ 15 minutes sur un PC muni d'un processeur pentium IV.

## 4.3.4 Cas d'étude

### 4.3.4.1 Site d'étude test

Le site d'étude test **D** sélectionné est d'une surface de  $3 \text{ km}^2$ . Il correspond un sous bassin versant du Bourdic incluant lui-même le bassin de Roujan et le site **B** utilisé précédemment. L'extension de ce site correspond aux limites du bassin test dont l'exutoire est marqué par un point bleu sur la figure 4.24, sauf dans les parties nord-est et sud de cette zone qui débordent du bassin versant. Pour cette zone, le lattice cadastral contient 1855 arêtes, 1348 noeuds (points verts). On peut visualiser le vrai réseau hydrographique artificiel du bassin test sur la droite de la figure 4.24.

### 4.3.4.2 Topologie du vrai réseau

L'épaisseur du trait et le niveau de gris du réseau représenté sur la droite de la figure 4.24 sont relatifs à l'ordre HS  $w$  de chaque arête du réseau. Ce réseau contient 239 arêtes, 240 noeuds dont 123 sources, 116 confluences et 1 exutoire. La matrice de ramification de ce réseau d'un bassin d'ordre  $w_{max} = 5$  est :

<sup>13</sup>chaque enregistrement de cette table décrit un arc, i.e. son identifiant, l'identifiant de ses noeuds terminaux et sa longueur.



1	2	3	4	5
0.486	0.514			
0.357	0.250	0.393		
0.500	0.083	0.167	0.250	
0.154	0.461	0.231	0.077	0.077

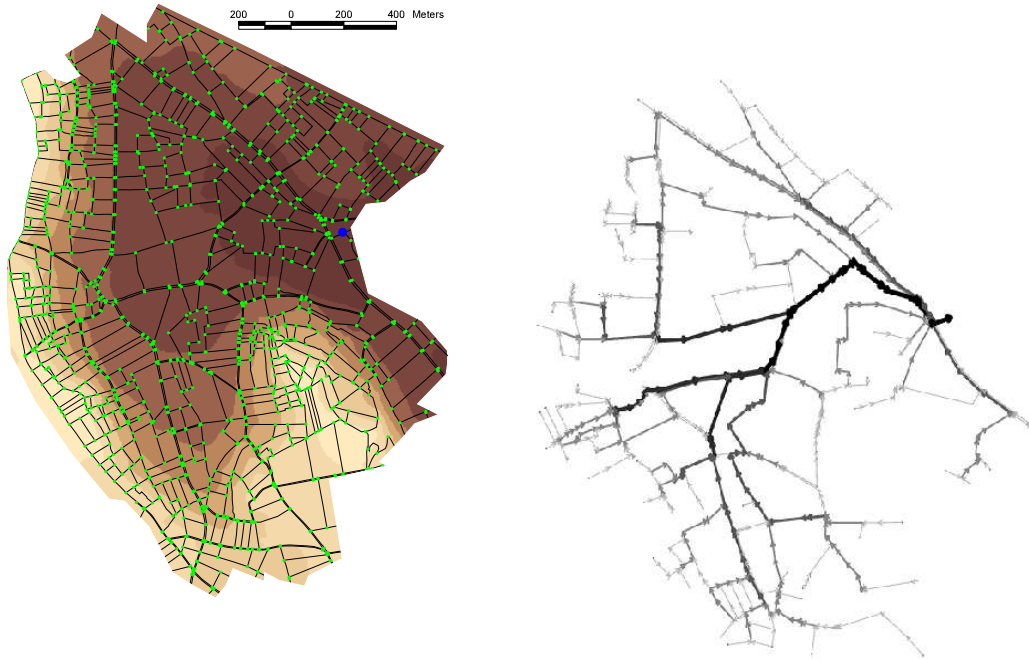


FIG. 4.24 – Site test **D**. Gauche : Lattice parcellaire sur M.N.T. LiDAR à 1 m (changement de couleur tous les 8 m allant de 58 m (brun sombre) à 128 m (brun clair) ; Droite : vrai réseau avec une épaisseur et un niveau de gris relatifs à l'ordre  $w$  de Horton-Strahler (bassin d'ordre  $w_{max}=5$ ).

Sur la figure 4.25, on représente l'ensemble des arêtes données provenant soit :

- de la télédétection de profils altimétriques LiDAR (méthode ondelettes, modèle 'moyen') "lissés" par arête (voir section précédente) (noir),
- de la BD-ORTHO, ©IGN sur les réseaux hydrographiques de la zone ou de l'interprétation des réserves foncières collectives créées au niveau du cadastre pour les collecteurs principaux (bleu),
- de l'interprétation sur un M.N.T des quelques ravines très marquées correspondant à des chemins encaissés, portions du réseau (rouge).

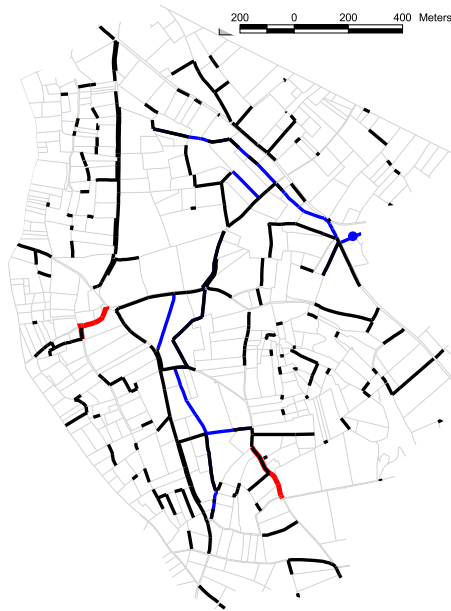


FIG. 4.25 – Ensemble des arêtes données sur le zone test pour conditionner les simulations de réseaux.

#### 4.3.4.3 Données pour l'orientation du Lattice

L'orientation du lattice par la topographie du terrain est réalisée par valuation des noeuds suivant une altitude. Cette altitude est calculée à partir d'un M.N.T LiDAR (Beauland, 2003) (Fig. 4.26). Ce M.N.T à maille carrée régulière de 1 m a été réalisé à partir des nuages de points de dernier retour d'onde du LiDAR (voir p. 174) suivant une procédure de traitement simple (altitude minimum par maille) et pour partie manuelle (adaptation de filtres morphologiques locaux de type ouverture) (Wiechert *et al.*, 2000). Sa qualité n'a été évaluée que sur zones planes et sans végétation (Saidouni, 2003) mais on le sait cependant relativement bruité (en  $z$  mais aussi en  $x,y$  (Bretar, 2006)) sur les zones de végétation et là où existent de fortes discontinuités du terrain, ce qui est souvent le cas en limite de parcelle (Fig. 4.26). Des travaux d'amélioration de la qualité de ce M.N.T sont en cours, notamment par classification du nuage de points LiDAR mais n'ont pu être, en temps voulu, insérés dans ce travail (Bretar, 2006).

La qualité altimétrique du M.N.T et les erreurs de positionnement planimétrique entre le M.N.T et le lattice cadastral confèrent à l'altitude des noeuds du lattice une forte incertitude.

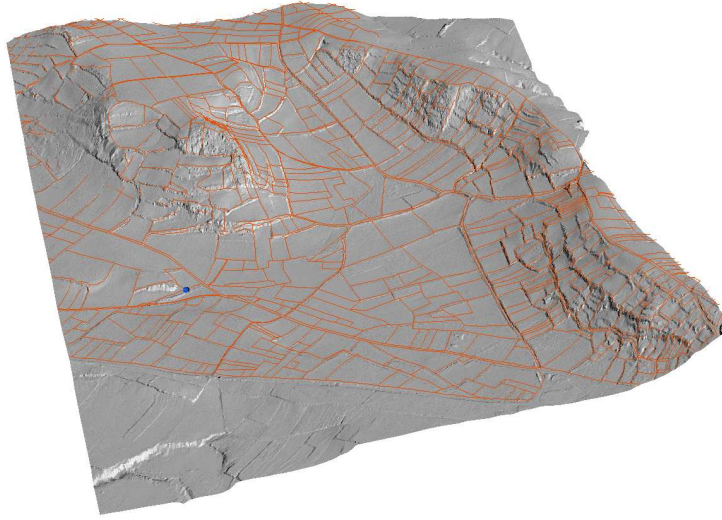


FIG. 4.26 – Bloc diagramme du M.N.T LiDAR (exagération verticale \* 5) sur le site test **D**. Le lattice parcellaire est plaqué en rouge sur ce M.N.T. L'exutoire du site test est en bleu.

## 4.3.5 Résultats

### 4.3.5.1 Ajout d'arêtes entre îlots parcellaires

La figure 4.27 représentent en rouge les arêtes ajoutées selon la procédure décrite auparavant avec un critère de distance  $\Delta h = 15$  m, largeur englobant la largeur maximale entre îlots parcellaires sur le site test. On peut voir sur cette figure des arêtes supplémentaires situées pour la plupart en travers des routes et chemins et très peu situées à l'intérieur des îlots.

### 4.3.5.2 Orientations du lattice

**4.3.5.2.1 Sensibilité à  $\Delta Z$**  L'orientation du lattice est sensible à l'incertitude sur l'altitude des noeuds du lattice. Sur la figure 4.28, on représente en rouge les noeuds (et arêtes) du lattice accessibles depuis l'exutoire lorsque l'on 'remonte' pas à pas le lattice depuis l'exutoire et ce, suivant des valeurs croissantes de désorientation  $\Delta Z$  : on réalise un équivalent de l'algorithme watershed (Soille, 1999) dans le lattice orienté en partant de l'exutoire. Une valeur  $\Delta Z = 1.25m$  permet d'atteindre la totalité du lattice à l'intérieur du bassin depuis l'exutoire (dont l'extension correspond à peu près aux limites du bassin versant de l'exutoire), mais dans ce cas on "déborde" également du bassin. Un équilibre entre accessibilité interne au bassin et non débordement du bassin est atteint pour  $\Delta Z = 0.75m$ . On retiendra par la suite cette valeur pour le paramètre  $\Delta Z$ .

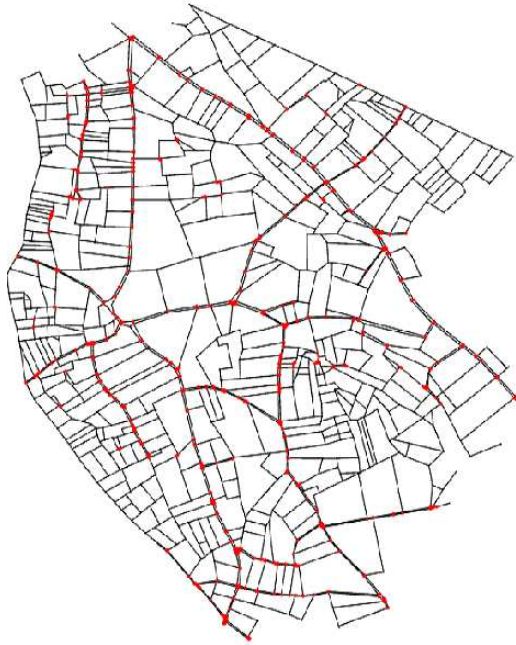


FIG. 4.27 – Arêtes ajoutées (en rouge) entre noeuds du lattice suivant un critère de distance (15m) entre ces noeuds.

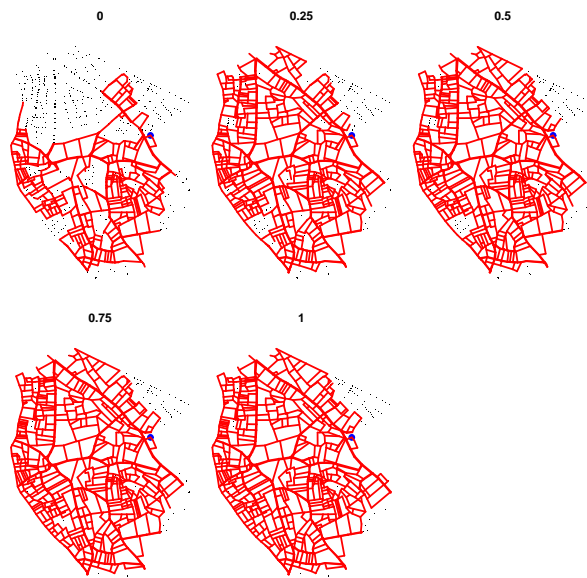


FIG. 4.28 – Noeuds accessibles par cheminement amont dans le lattice orienté depuis l'exutoire (bleu) pour des valeurs croissantes de  $\Delta_Z$

**4.3.5.2.2 Degrés inférieurs des noeuds du lattice orienté** Les degrés inférieurs du lattice orienté (voir leur définition sur la figure 2.4 page 36), suivant les valeurs du paramètre  $\Delta Z$ , conditionnent les possibilités de tirage aléatoire d'un noeud aval lors des cheminements amont-aval : seuls les noeuds de degré inférieur ou égal à 2 pourront être support d'un processus aléatoire (choix aléatoire d'un seul noeud aval). La distribution des degrés inférieurs des noeuds du lattice représente donc les degrés de liberté (choix possibles) pour ces cheminements aléatoires amont-aval.

Afin de quantifier ces degrés de liberté pour les simulations, on calcule les degrés inférieurs des noeuds du lattice suivant différentes valeurs de  $\Delta Z$ . La figure 4.29 représente à gauche tous les noeuds du lattice avec un code couleur correspondant à leur degré inférieur calculé pour  $\Delta Z = 0$  : du jaune très clair (0) au rouge (5). Les degrés supérieurs à 3 sont peu nombreux et résultent de l'ajout d'arêtes en travers des routes et chemins. A droite, on représente l'évolution du degré inférieur moyen des noeuds du lattice pour des valeurs croissantes de  $\Delta Z$  suivant un pas croissant de 0.25 m (de 0 à 12 m). On peut observer, qu'à partir de  $\Delta Z = 0.25m$  (droite rouge horizontale), on a en moyenne un degré inférieur moyen de 2 (deux noeuds avals accessibles par noeud du lattice) et, pour  $\Delta Z = 0.75m$  (droite bleue verticale), le degré inférieur moyen est d'environ 2.5. L'asymptote qui représente le degré moyen des noeuds du lattice non orienté est d'environ 3.5.

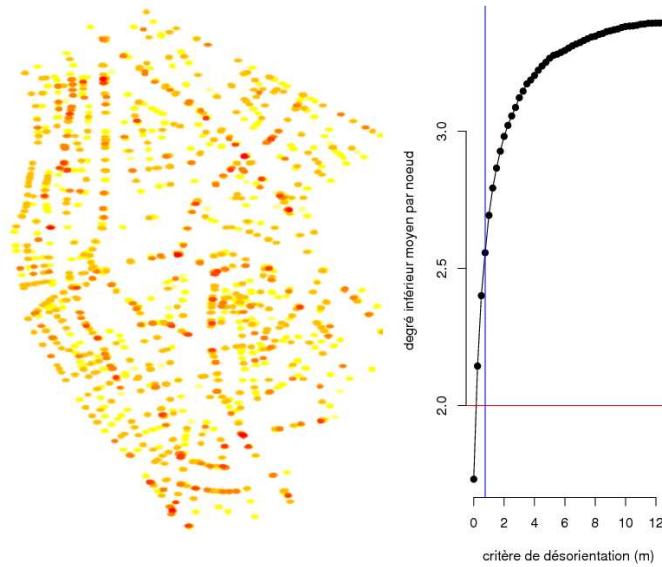


FIG. 4.29 – Gauche : degrés inférieurs des noeuds du lattice pour  $\Delta Z = 0$  (jaune clair  $d_- = 0$  à rouge  $d_- = 4$ ); Droite : Evolution du degré inférieur moyen des noeuds du lattice en fonction de  $\Delta Z$

### 4.3.5.3 Distribution spatiale des sources

Sur le site d'étude, les fossés ont un rôle d'assainissement agricole, notamment dans les zones de contacts entre couches géologiques de perméabilité différentes qui créent alors des "mouillères". Ces zones sont, sur le site d'étude, plutôt situées en tête de réseau au pied des ruptures de pente, convexes ou concaves. Si l'on observe globalement la distribution des sources dans l'espace du bassin test sur la figure 4.30 (points bleus), on observe une hétérogénéité de leur densité dans l'espace. Ces hétérogénéités sont en partie expliquées par la dimension des mailles du lattice cadastral et la densité des noeuds du lattice résultante (points verts). On observe également que les sources semblent plus regroupées le long de certaines courbes de niveaux représentées en rouge (83, 101 et 118 m). Ces altitudes correspondent aux affleurements en surface de couches géologiques de perméabilités différentes, hydromorphes, situées sur les ruptures de pente des zones de coteaux et de glacis (Fig. 4.30). Ces densités apparentes élevées pour ces altitudes se confirment, une fois annulé l'effet de la densité des noeuds du lattice, comme le montre l'histogramme à droite de la figure 4.30 : cet histogramme comptabilise par tranche d'altitude de 5m la proportion de sources dans les noeuds du lattice. On modélise cette distribution spatiale suivant une altitude des sources du réseau.

Une modélisation par processus ponctuel spatial des sources dans l'espace n'est, dans ce cas, pas adaptée car les effets spatiaux apparents sont essentiellement dus à la taille des mailles du lattice et la forme de l'espace support du processus, les noeuds du lattice, est complexe. En revanche, on considère que pour un bassin sédimentaire comme celui de la Payne, avec des affleurements créant des mouillères suivant des altitudes successives constantes, les sources, contraintes par la structure du lattice (noeuds), se concentrent de manière gaussienne autour de ces altitudes.

On modélise donc la distribution des sources par une combinaison de lois normales (courbes rouges sur la figure 4.30) de même amplitude maximum (16 %) et de même variance (2.25) mais centrée sur des valeurs d'altitude différentes (83, 101 et 118 m). Le modèle s'écrit, pour cette proportion de sources, fonction de l'altitude  $z$ , pour  $n$  altitudes centrales :

$$prop_S(z) = \theta_1 * \left\{ \sum_{i \in \{1, \dots, n\}} f(\mu_i, \sigma_1) \right\}$$

avec  $f(\mu, \sigma)$ , fonction de densité d'une loi gaussienne de paramètres  $\mu, \sigma$ , moyenne et écart-type.

Les paramètres ajustés sont :  $\hat{\theta}_1 = 2.25$ ,  $(\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3) = (83, 101, 118)$ ,  $\hat{\sigma}_1 = 6$ .

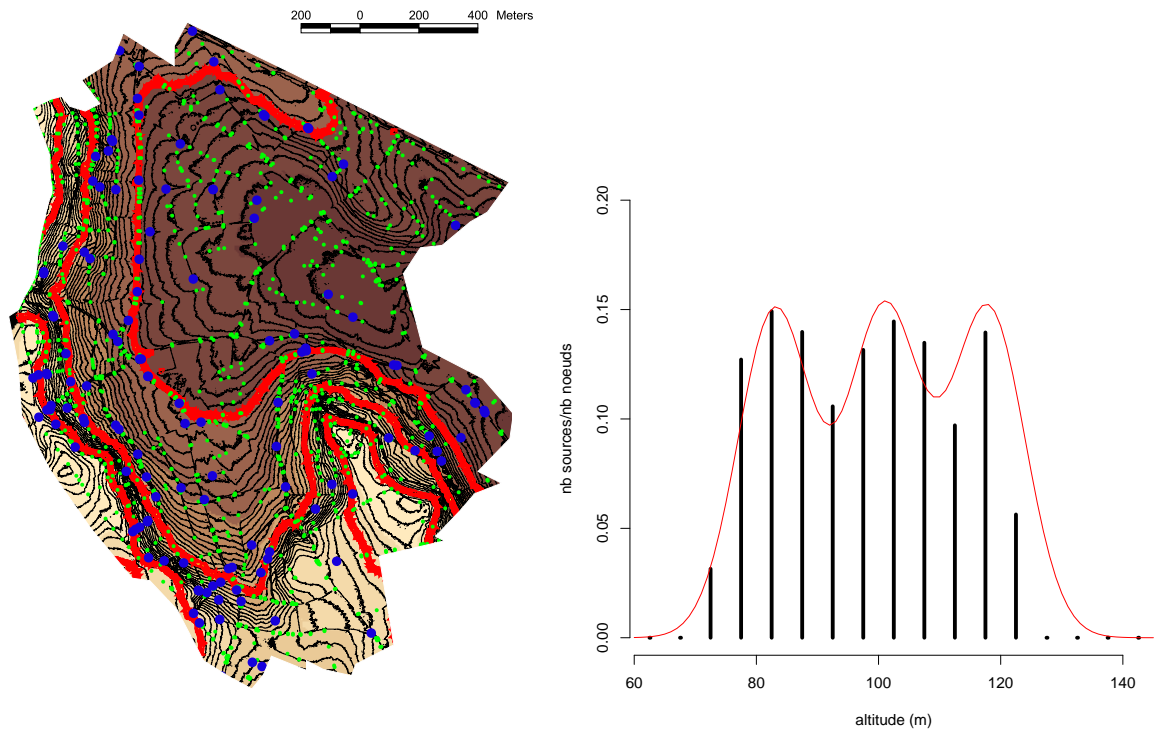


FIG. 4.30 – Gauche : Distribution des sources (points bleus) du vrai réseau hydrographique artificiel et des noeuds (petits points verts) du lattice (contours iso-altitude en noir tous les m) sur le bassin test **D** ; Droite : Rapport entre le nombre de sources et le nombre de noeuds du lattice par tranche d'altitude (tous les 5 m) et lois normales ajustées (rouge).

#### 4.3.5.4 Distribution d'un ensemble de simulations

Suivant la méthode décrite auparavant, on réalise 100 simulations. Ces simulations sont réalisées suivant une règle de tirage équiprobable des noeuds "avals" et suivant un paramètre de désorientation du lattice  $\Delta Z = 0.75m^{14}$ . A titre d'exemple, la figure 4.31 montre 11 arborescences simulées avec une représentation de l'ordre de Horton-Strahler par arête de ces arborescences. On étudie ensuite la distribution de ces arborescences simulées selon des critères topologiques (comparaison de matrices de ramification), géométriques (comparaison de densités de drainage globales et locales), et topographiques (comparaison des distributions des pentes du réseau), critères reliés au fonctionnement hydrologique de ces réseaux.

<sup>14</sup>le paramètre de rehaussement maximal des puits est le même que  $\Delta Z$ , soit 0.75 m et le paramètre  $R_{max}$  utilisé à été fixé à 10.

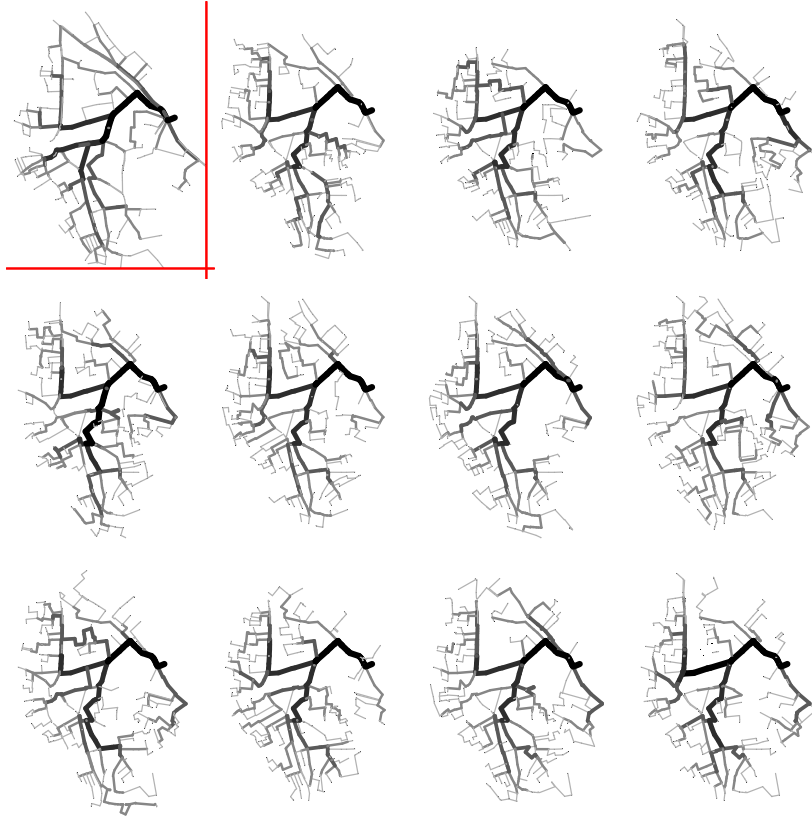


FIG. 4.31 – 11 exemples de simulations conditionnelles de réseaux comparés au vrai réseau (en haut à gauche) avec une épaisseur de trait et un niveau de gris relatifs à l'ordre de Horton-Strahler  $w$ , allant de 1 à 5.

**4.3.5.4.1 Critères topologiques** Par construction, la distribution des sources des arborescences simulées, premier critère topologique, respecte un modèle donné (selon l'altitude et la densité de noeuds du lattice).

Un deuxième critère topologique étudié est la structure de la matrice de ramification. Pour chaque matrice de ramification d'un réseau simulé, on calcule son vecteur de distance à la matrice de ramification du vrai réseau. Pour ce faire, on calcule la différence, élément par élément  $p_{w,i}$ , des matrices suivant un ordre de bassin  $w = 5$ . On ne calcule pas cette différence pour les éléments de matrice  $p_{w-1,w-1}$ , car elles s'expriment comme une combinaison linéaire des autres  $p_{w,i}$ . On étudie ensuite la distribution de cette distance élément par élément (0 représente la valeur du vrai réseau par élément  $p_{w,i}$ ). Sur la figure 4.32, on représente par boîtes à moustaches ces distributions calculées sur les 100 réseaux simulés. Sur cette figure, les barres horizontales rouges représentent les quantiles 2.5% et 97.5% de ces simulations. La barre bleue horizontale, à 0, représente le vrai réseau. La barre verte est la valeur  $p_{w,i}$  de la matrice asymptotique d'un arbre aléatoire (Penaud, 1988).



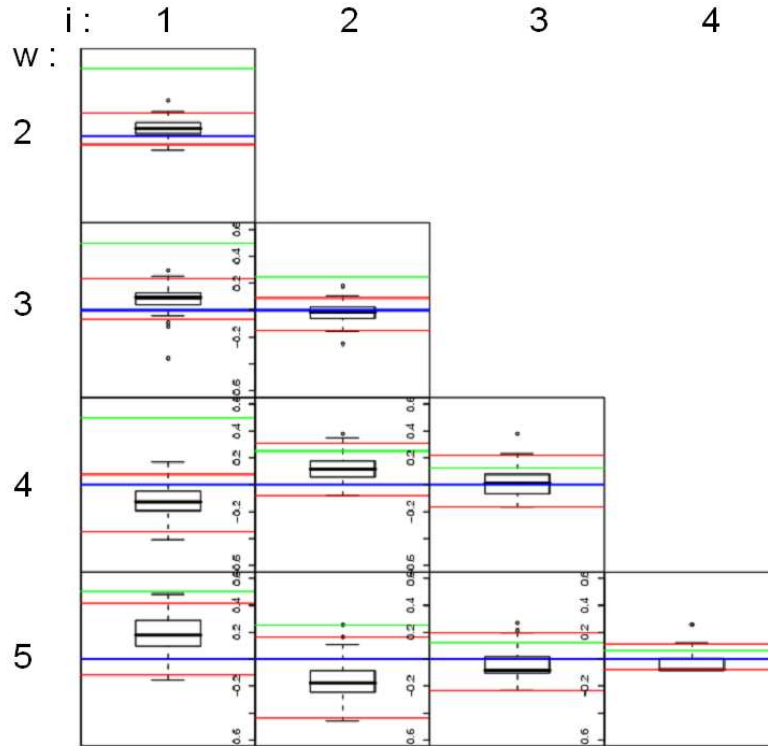


FIG. 4.32 – Boîtes à moustaches de la différence, élément par élément  $p_{w,i}$ , entre les matrices de ramification des réseaux simulés et la matrice de ramification du vrai réseau pour un ordre de bassin  $w = 5$ ; seuls les éléments  $p_{w,i}$  avec  $i < w$  sont représentés.

Pour chaque élément de ces matrices, le vrai réseau est contenu dans l’enveloppe à 95 % des simulations : il n’est donc pas significativement différent sur ce critère topologique. Cependant certaines distributions de  $p_{w,i}$  des réseaux simulés ne sont pas toujours centrées sur celle du vrai réseau, notamment pour les bi-ordres  $p_{3,1}$ ,  $p_{4,2}$ ,  $p_{5,1}$  et  $p_{5,2}$ .

En revanche, sur ce critère de matrice de ramification, on peut remarquer que les topologies des réseaux, simulés et vrai, sont significativement différentes de celle d’un réseau ‘purement aléatoire’ (lignes horizontales vertes), i.e. sans aucune contrainte de topographie et topologie.

**4.3.5.4.2 Critères géométriques** La figure 4.33 représente l’histogramme du linéaire total des réseaux simulés. Le linéaire total du vrai réseau, 26580 m, est bien au centre de la distribution (écart de -84 m entre la moyenne pour les réseaux simulés et le vrai réseau, soit un très faible biais de -0.3 pour mille). Les linéaires totaux simulés ont un écart-type qui représente 5.7 % du vrai linéaire total. Si une règle de tirage des noeuds aval favorisant les plus courts chemins lors des chemine-ments aléatoires amont-aval entre arêtes détectées avait été appliquée, on obtiendrait certainement par construction un linéaire total simulé plus faible, ce qui n’est pas forcément souhaitable.

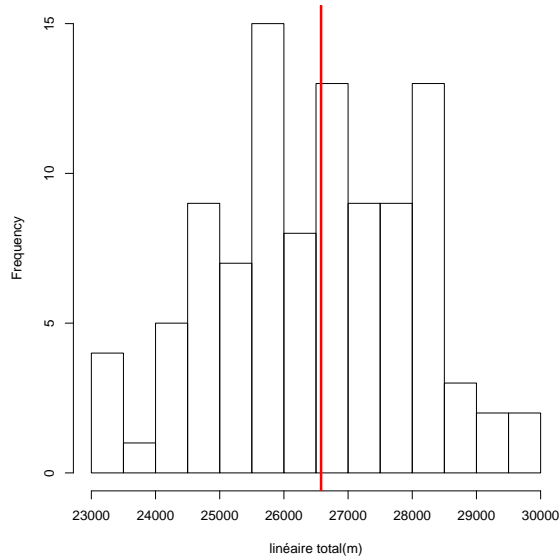


FIG. 4.33 – Histogramme du linéaire total par réseau simulé. Le linéaire total du vrai réseau est la ligne rouge.

On calcule ensuite une "densité de drainage" (non classique car égale au rapport entre linéaire du réseau et linéaire des arêtes du lattice) par sous-secteur du bassin test.

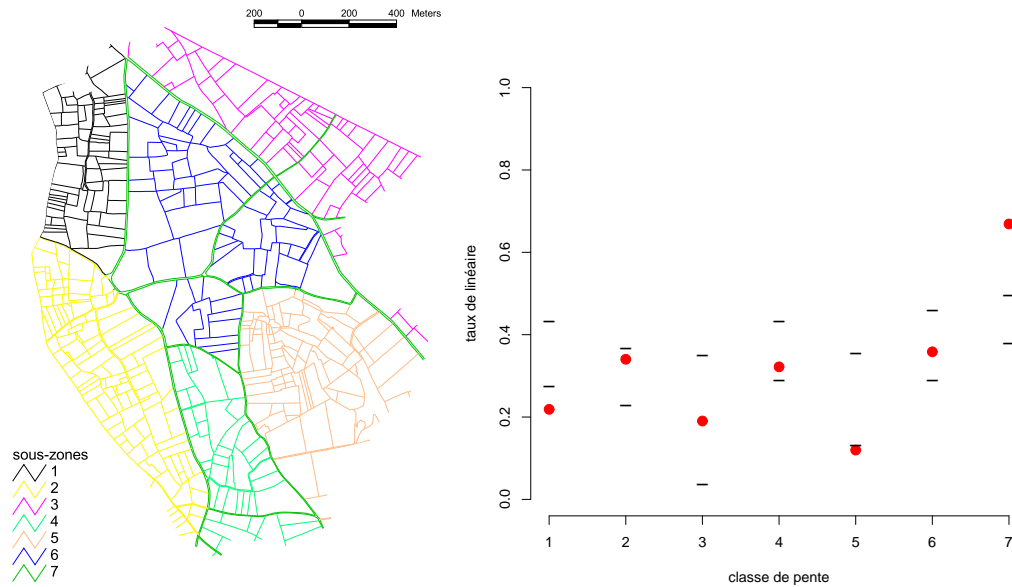


FIG. 4.34 – Gauche : segmentation du lattice en 7 sous-secteurs ; Droite : densité de drainage sur le lattice pour le vrai réseau (points rouges) et quantiles 2.5% et 97.5% (traits noirs) pour les réseaux simulés.

Ces sous-secteurs, choisis arbitrairement, distinguent les versants suivant leur orientation ainsi que les bords de route regroupés au sein d'un même sous-secteur. Ces sous-secteurs sont représentés par code couleur à gauche de la figure 4.34. Globalement, en comparant les réseaux simulés au réseau vrai, l'ordre des densités de drainage entre sous-secteurs est respecté. En revanche, pour certains de ces sous-secteurs, on obtient une densité de drainage du vrai réseau significativement différente de celles simulées (droite de la figure 4.34) : on a une surestimation sur les sous-secteurs 1 (terrasses au nord-ouest) et 5 (plateau au sud-est) et une sous-estimation du secteur 7 (bords de route).

**4.3.5.4.3 Critères topographiques** On s'intéresse à la distribution des pentes des arêtes des arborescences simulées. La pente est calculée, pour chaque arête de l'arborescence comme le rapport entre la dénivellée positive de ses deux noeuds et sa longueur. Les classes de pente choisies sont délimitées arbitrairement par les deciles des pentes des arêtes du lattice (indépendamment de leur linéaire). Les bornes inférieures des 10 classes qui en découlent sont, exprimées en % : 0.0, 0.5, 1.0, 1.6, 2.5, 3.4, 5, 7.2, 11.0, 17.0. Pour chaque arborescence simulée, on calcule le linéaire total de ses arêtes par classe de pente. Sur la figure 4.35, on représente l'enveloppe à 95% bilatérale de ces linéaires totaux (en m) par classe de pente croissante. La courbe rouge représente la distribution des pentes d'arêtes sur le vrai réseau. Sur ce critère, le vrai réseau est globalement contenu à l'intérieur de l'enveloppe mais est il situé en limite de l'enveloppe pour certaines classes, notamment pour les pentes fortes : cet écart peut être relié à l'impact des sur-détections LiDAR sur les zones de terrassettes, contenant ces fortes pentes.

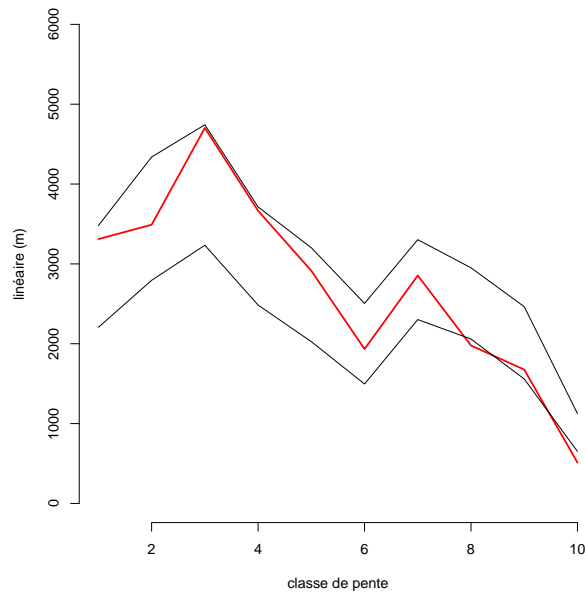


FIG. 4.35 – Distribution du linéaire selon 10 classes de pente : enveloppe de confiance à 95% bilatérale pour les réseaux simulés et le vrai réseau (ligne rouge).

### 4.3.6 En conclusion

Dans cette section, nous avons proposé un algorithme stochastique qui, pour un exutoire donné dans un lattice orienté par un M.N.T et à partir d'un ensemble d'arêtes de réseau données, d'un modèle de distribution des sources dans l'espace des noeuds du lattice, simule des arborescences. Ces arborescences, de même exutoire, passent par la plupart des arêtes données et respectent une loi de distribution spatiale des sources. Le caractère stochastique de l'algorithme vient d'un cheminement aléatoire amont-aval dans le lattice vers l'exutoire et de la réalisation d'une loi probabiliste de distribution topographique, donc spatiale, des sources. La loi de distribution topographique des sources utilisée étant facile à simuler (mélange de lois gaussiennes), une simple réalisation de cette loi a été utilisée pour chaque simulation. Pour des distributions plus complexes et dans le cadre d'un raffinement de la méthode de simulation (distributions topographiques conjointes des sources et des confluences, . . .) d'autres algorithmes pourraient être envisagés, notamment un algorithme de type MCMC (Metropolis-Hastings). La difficulté, dans ce cas, serait plus liée aux conditions initiales de l'algorithme, fortement dépendantes des segments connus, i.e. télédétectés, qui, combinées aux nombreuses contraintes de l'espace support des simulations, pourraient le rendre peu efficace (beaucoup d'itérations nécessaires pour converger vers la distribution souhaitée).

Quoiqu'il en soit, avec l'algorithme proposé et le conditionnement provenant des résultats de télédétection LiDAR des fossés, on reproduit une distribution de réseaux réaliste selon des critères topologiques et topographiques. Ces réseaux simulés présentent cependant quelques écarts par rapport au vrai réseau, notamment selon des critères de géométrie locale, laquelle géométrie locale est assez spécifique au site test d'étude. Par ailleurs, un seul jeu de paramètres de l'algorithme proposé n'a été utilisé. Nous n'avons donc pas mesuré le poids de ces paramètres (critère de désorientation du lattice, règle de tirage d'un noeud aval. . .) sur le réalisme des simulations. Des études de sensibilité des simulations pour différentes valeurs de ces paramètres restent à réaliser.

La méthode proposée est généralisable à d'autres bassins, à d'autres milieux, sous réserve de pouvoir modéliser une distribution spatiale des sources, d'orienter le lattice par un M.N.T et de conditionner la simulation par un ensemble d'arêtes données, notamment l'ossature principale du vrai réseau.

## 4.4 Bilan

Dans ce chapitre, nous avons proposé une méthodologie de simulation conditionnelle de réseaux arborescents dans un lattice orienté. Cette méthodologie se déroule en deux phases.

La première phase est métrologique. On teste le pouvoir de détection des fossés qui composent ce réseau sur des données topographiques de télédétection aéroportées à très haute définition, les données LiDAR. Pour ce faire, une méthode originale de détection ponctuelle des fossés sur le lattice parcellaire a été développée. Cette méthode s'appuie sur la classification de profils altimétriques LiDAR suivant leur "forme", notamment les discontinuités qu'ils contiennent marquant la présence de fossés. Elle pourrait être appliquée pour la détection ou la caractérisation d'autres éléments linéaires des paysages cultivés influents sur le régime et la qualité des eaux et formant, de même, des discontinuités altimétriques, situées sur le lattice des limites cadastrales : haies, talus, etc.

Cependant, même après avoir filtré les résultats de détection ponctuels, le taux de détection de fossés obtenu reste assez faible, ce qui en soit n'a rien d'étonnant compte tenu du contexte du site d'étude. Ce taux ne semble pas dû principalement à une mauvaise performance des méthodes mises en oeuvre, de classifications comme d'analyses de forme : on obtient des performances similaires pour différentes méthodes traduisant la forme des profils, par transformée en ondelettes ou morphologie mathématique (watershed), des performances globales similaires pour différentes méthodes de classification (tests de classification par la méthode des forêts aléatoires en annexe p. 189), performances toutes comparables à une classification visuelle, non automatique. Cette faible détection est plus à relier à la nature du réseau sur le site test (réseau et morphologie complexes, réseau diversement entretenu et souvent végétalisé, c.f photos du réseau page 17) et à la nature des données LiDAR utilisées pour représenter le terrain et non la surface lorsque l'acquisition est réalisée en période de végétation active. Pour autant, nous n'excluons pas la possibilité de rendre les méthodes plus performantes, par exemple en améliorant la représentation du terrain dans les profils altimétriques par classification préalable de type sol-sursol des points LiDAR (Bretar, 2006).

Quoiqu'il en soit, on peut surtout attendre des progrès significatifs dans le taux de détection des fossés par utilisation de données LiDAR hors période de végétation ou suivant d'autres modes d'échantillonnages spatiaux par le LiDAR.

La deuxième phase est une phase d'interpolation algorithmique. On connecte au travers du lattice cadastral les arêtes de réseau détectées, ou connues par suivant des bases de données facilement accessibles, pour former une arborescence. Cette connection s'opère selon des cheminements aléatoires dans le réseau qui suivent des critères d'altitude. On complète ensuite cet ensemble par des branches aléatoires pour converger vers une distribution spatiale (topographiques) des sources du réseau donnée. L'analyse des résultats de ces simulations conditionnelles suivant des critères topographiques, géométriques et topologiques, montre que par ces moyens, on mime des réseaux à l'architecture réaliste. Ce résultat tempère la faible détection quantitative du linéaire de fossés par LiDAR et laisse penser que cette détection, couplée aux autres données facilement accessibles, peut être suffisante en représentant surtout les artères principales du réseau.

Cependant plusieurs questions n'ont pas été traitées et restent en suspens. Tout d'abord, on peut se demander si le réalisme descriptif obtenu se traduit également par un réalisme dans le fonctionnement hydrologique. Autrement dit, les incertitudes de spatialisation résultant de cette démarche ont-elles une influence significative lorsqu'elles sont propagées dans un modèle hydrologique? A l'inverse, peut-on se permettre d'être moins précis dans la procédure de spatialisation? Des études de sensibilité de modèles hydrologiques utilisant ces premiers résultats doivent être menées pour répondre à cette question.

Par ailleurs, nous n'avons pas quantifié réellement quel était le poids des données LiDAR et celui des méthodes d'analyse dans les résultats obtenus. L'emploi des données LiDAR a été double : 1- les données "brutes" (points topographiques irréguliers) sont utilisées pour détecter ponctuellement les fossés dans le lattice cadastral ; 2- un M.N.T réalisé à partir de ces mêmes données LiDAR est utilisé pour orienter le lattice dans l'algorithme de simulation d'une arborescence. La démarche proposée reste sur le principe généralisable avec d'autres données. Ces autres données peuvent être d'autres M.N.T utilisés pour orienter le lattice cadastral et toute source de données donnant une information partielle sur la localisation du réseau dans le domaine étudié. Par exemple, la méthodologie proposée en sous-section 4.2.1 peut également être appliquée sur des profils radiométriques d'images à très haute résolution spatiale dans la mesure où les discontinuités topographiques recherchées créent également des discontinuités de radiométries (ombres). Quelques tests réalisés sur le site d'étude montrent un pouvoir de détection des fossés qui peut être intéressant sur ces images, restant cependant inférieur au pouvoir de détection LiDAR dans des conditions comparables (annexe p. 194) (Bailly *et al.*, 2006a). Pour l'orientation du lattice, des simulations conditionnelles d'arborescences à partir du lattice cadastral orienté par d'autres M.N.T., à maille plus large, moins précis, pourraient être menées. On pourrait alors évaluer le poids des données LiDAR en comparant le réalisme des simulations utilisant ces autres données.

De même, avec la méthodologie de simulation conditionnelle proposée, on est désormais en mesure de tester la sensibilité des réseaux simulés en utilisant différentes méthodes d'analyse et de classification des données LiDAR (sélection de tel ou tel arbre de classification, analyse de formes par ondelette ou autre, classification par forêts aléatoires ou d'autres méthodes, etc.). On pourrait ainsi, par comparaison, juger plus avant des performances ou biais de ces méthodes pour spatialiser des réseaux hydrographiques artificiels de paysages cultivés et juger plus objectivement des conséquences de la sur-détection ou sous-détection de fossés.



# Conclusions et perspectives

## *Synthèse des résultats*

En considérant que l'hétérogénéité spatiale des réseaux hydrographiques artificiels conditionne fortement les transferts hydrologiques des paysages cultivés, l'objectif de ce travail de recherche était de proposer des méthodologies qui permettent de représenter ces réseaux dans l'espace à partir de rares données échantillonnées sur le terrain.

Pour répondre à cet objectif, des méthodes de spatialisation de ces réseaux ont été développées selon deux angles de vue de leur hétérogénéité : la spatialisation du réseau lui-même, de son architecture et la spatialisation de propriétés locales le long du réseau. Toutes les méthodes développées s'inscrivent dans une démarche relativement classique de simulation stochastique qui met en synergie échantillonnages de terrain, modèles d'organisation spatiale et données de télédétection. L'originalité des méthodes proposées est davantage liée à la nature atypique des objets spatiaux étudiés : ces méthodes sont adaptées aux objets linéaires organisés en réseaux arborescents et orientés, et pour certaines, elles sont adaptées à des hétérogénéités spatiales sous forte influence anthropique.

Pour la simulation de propriétés le long du réseau, deux modèles probabilistes spatiaux ont d'abord été proposés : une fonction spatiale aléatoire (géostatistique), orientée de l'amont vers l'aval, et un processus ponctuel aléatoire sur le réseau qui le découpe en segments homogènes. Pour ces deux modèles, c'est principalement l'introduction d'hypothèses additionnelles aux concepts de base des méthodes probabilistes spatiales usuelles, hypothèses d'indépendance liées à l'orientation ou à la topologie du réseau, qui permet leur définition sur des réseaux arborescents orientés. Dans le cas géostatistique, l'adaptation, pour des raisons topologiques, d'une technique de simulation conditionnelle existante suivant une séquence des calculs numériques aval-amont dans le réseau permet la simulation de valeurs tout le long du réseau.

Pour la spatialisation de l'architecture du réseau sous forme d'arborescence orientée, de l'amont vers l'aval, une méthode de simulation par connexion, interpolation stochastique de segments de réseaux déterminés par télédétection au travers d'un lattice parcellaire (réseau maillé des limites de parcelles) orienté par la topographie, a été élaborée. L'approche proposée n'ayant pu s'appuyer sur des études pré-existantes aux objectifs similaires, elle est allée depuis l'exploration de méthodes de prises d'information par télédétection jusqu'au développement de méthodes de simulation du réseau à partir d'échantillonnages partiels sur le réseau. Elle s'est donc décomposée en



deux phases successives. La première, métrologique, cherche à détecter les fossés qui composent le réseau à partir de données topographiques acquises par télémétrie laser aéroportée. La deuxième est algorithmique et conceptuellement indépendante de la phase précédente : elle complète les segments de réseaux détectés, ou connus suivant d'autres sources de données, par des segments du lattice parcellaire choisis aléatoirement mais sous plusieurs contraintes : contraintes topologiques et topographiques. Les contraintes topographiques sont liées à l'utilisation d'un modèle numérique de terrain, acquis également par télédétection. Du point de vue des résultats, le taux de succès de la détection automatique par LiDAR des fossés est faible mais à priori conforme à celui attendu compte tenu de la nature du terrain d'étude et des données expérimentées. Cette faible détection semble cependant suffisante puisque l'architecture des réseaux simulés à partir de ces ensembles connus, détectés, semble réaliste. Le réalisme de ces simulations a été jugé suivant des critères de tests statistiques réalisés sur des index de géométrie et de topologie du réseau.

Plusieurs résultats des méthodes proposées sont inhérents à la démarche choisie :

- les simulations de propriétés locales du réseau ou de son architecture pouvant être conditionnées par des échantillons sur le réseau, on peut les utiliser d'une manière générale pour compléter les données manquantes ; pour les propriétés locales le long du réseau, on peut même les utiliser en l'absence totale de données (simulations non conditionnées par des données).
- les incertitudes de spatialisation sont représentées au travers de plusieurs réalisations de ces simulations stochastiques. Ces méthodes s'apparentent à la famille générale des méthodes de Monte-Carlo, mais spatiales.
- les développements méthodologiques proposés produisent des simulations réalistes suivant des critères de structure spatiale mais aussi de distribution, notamment par rapport aux queues de distribution des propriétés locales qui peuvent être déterminantes sur le fonctionnement des hydro-systèmes.

En revanche, un résultat peut-être moins attendu est que les simulations de l'architecture du réseau ou de ses propriétés locales apparaissent également fortement conditionnées par la structure de l'espace qui supporte ces simulations : une arborescence dans le cas de propriétés locales et un lattice orienté dans le cas de l'architecture du réseau. Ceci laisse penser que cette contrainte "structurelle" de l'espace support permettrait de simuler des réseaux réalistes pour un taux d'échantillonnage assez faible sur le réseau.

Cependant, pour les hétérogénéités explorées, nous sommes restés dans un cadre de calibration des méthodes, en ne travaillant que sur un seul réseau. Juger plus avant du réalisme des simulations obtenues par ces méthodes, suivant différents taux d'échantillonnages nécessiterait, à terme, de les tester sur de nouveaux réseaux.

Une autre limite de cette recherche, du moins pour son application aux réseaux hydrographiques artificiels, est qu'une seule propriété locale le long du réseau, la largeur, n'ait été étudiée. D'autres variables locales des réseaux artificiels d'intérêt hydrologique n'ont pas été explorées, notamment celles liées à la présence de végétation. Le travail d'analyse, de modélisation, d'évaluation potentielle par télédétection puis de simulation de ces propriétés reste à faire. Cependant, pour ces variables, reliées aux itinéraires techniques des parcelles cultivées environnantes, les contraintes appliquées aux simulations peuvent être envisagées différemment. Plutôt que des

contraintes issues d'un échantillonnage dense sur le terrain, on pourrait d'abord identifier les déterminants agronomiques de ces variables (type d'exploitation...), par enquêtes d'acteurs (Biarnes *et al.*, 2004). On pourrait ensuite voir comment ces déterminants influent sur l'organisation spatiale, et même spatio-temporelle, de ces variables. Quoiqu'il en soit, plus que les résultats des modèles spatiaux présentés (fonctions et paramètres), c'est surtout la démarche d'adaptation des modèles au type de domaine spatial étudié et aux hypothèses de processus physiques sous-jacents à la variable qu'il faut retenir. Pour de nouvelles variables, l'emploi d'une fonction spatiale ou d'un processus ponctuel aléatoires, voire des deux conjointement, est à relier au type de processus physique, naturel (hypothèse sous-jacente de variabilités continues) ou artificiel (hypothèse de pré-éminence d'unités homogènes d'actions anthropiques) à l'origine des variabilités étudiées.

### *Perspectives*

#### *Transposition des méthodes développées*

Les méthodes développées peuvent a priori se transposer à d'autres types de réseaux et propriétés locales sous influence combinée naturelle et anthropique. Par exemple, comme nous l'avons déjà évoqué en fin de chapitre 3, un modèle plus général adapté à des variabilités semi-anthropiques ou semi-naturelles le long du réseau serait un mélange des modèles processus ponctuel et géostatistique. Pour la spatialisation de propriétés locales le long du réseau, la méthode de simulation géostatistique peut être appliquée à d'autres variables supportées par d'autres types de réseaux hydrographiques, voire d'autres arborescences dès lors que l'orientation du réseau est à la base du processus physique sous-jacent de cette variable. Pour exemple, la méthode géostatistique proposée a été appliquée à une variable pédologique liée à un flux hydraulique amont-aval, la largeur d'un fluvisol, i.e. la zone d'épandage des crues, sur un réseau hydrographique naturel à l'échelle d'un bassin de quelques centaines de  $km^2$  (Monestiez *et al.*, 2005). La méthode de simulation par processus ponctuel, si elle paraît simple à mettre en oeuvre et peu sensible à l'orientation amont-aval dans le réseau, est très spécifique, car uniquement adaptée aux variabilités structurées en segments homogènes aléatoires. Un caractère moins aléatoire de la segmentation du réseau rendrait une démarche de type processus ponctuel dans une arborescence, difficile voire impossible du point de vue théorique.

D'un point de vue plus général, le champ de recherche sur la simulation spatiale d'objets artificiels des paysages cultivés est encore large ; cette recherche est peut-être un des premiers exemples. Une démarche similaire pourrait être entreprise dans le cas de spatialisation de propriétés de parcelles difficilement accessibles par télédétection (modes de conduite de l'entretien du sol, etc.), ou encore dans le cas des réseaux de haies, objets ou propriétés également influents sur le fonctionnement des paysages de bassins cultivés.

### *Combinaison des méthodes développées*

Les méthodes de simulation de réseaux présentées, de leur architecture ou de leurs propriétés locales ont été traitées indépendamment mais leur combinaison est bien sûr envisageable. Tout d'abord, comme nous l'avons déjà évoqué précédemment, une combinaison entre modèles géostatistiques et processus ponctuels pourrait être étudiée. A un niveau encore plus général, à partir d'un lattice cadastral, de données LiDAR, d'un modèle de structure spatiale des propriétés locales le long du réseau, voire de quelques observations en quelques sites du réseau, il est possible de simuler une arborescence complète : sa géométrie, sa topologie puis ses propriétés locales (sa largeur pour l'instant). Puisque les simulations proposées sont fortement conditionnées par la structure de l'espace qui les supporte, on pourrait ainsi étudier des distributions de propriétés de ces réseaux pour différents milieux, c'est à dire différentes morphologies et différentes structures parcellaires, puis en déduire des indicateurs synthétiques. On pourrait ainsi construire des typologies de bassins versants élémentaires suivant ces indicateurs, lesquelles typologies pourraient être utiles à des démarches de diagnostic de territoires pour de grandes étendues.

### *Utilisation hydrologique à court terme des méthodes développées*

L'ensemble du travail réalisé est resté volontairement dans un cadre descriptif de représentation spatiale de réseaux hydrographiques artificiels. Mais quelle perspective prioritaire donner à ce travail ? Doit-on aller vers encore plus de réalisme des simulations de réseaux hydrographiques artificiels par raffinement des méthodes de simulation ?

Il nous semble, au contraire, préférable de tester en premier lieu la base méthodologique proposée en la confrontant à ses utilisations potentielles : 1- sur d'autres réseaux et 2- dans des démarches d'évaluation, de prédiction du fonctionnement des paysages cultivés. Cette utilisation permettrait de juger des inflexions nécessaires, des développements inutiles en terme de modélisation des hétérogénéités spatiales, en terme de prise d'informations sur le terrain ou par télédétection.

On peut envisager cette utilisation des méthodes proposées pour différentes phases de construction ou d'utilisation de modélisations du fonctionnement des paysages cultivés. Ces phases, pour reprendre la taxonomie des modèles d'aménagement du territoire proposée par Steinitz (1990), sont surtout à court terme des phases de développement de modèles de processus des bassins versants cultivés.

En effet, on peut les utiliser pour accompagner les changements d'échelle spatiale de cette modélisation hydrologique et la hiérarchisation des processus et hétérogénéités "moteurs" à différentes échelles : application du modèle sur de plus vastes étendues (agrandissement spatial), sur de multiples sites (répétition spatiale) et confrontation d'observations hydrologiques et de sorties de modèles hydrologiques. Elles peuvent ainsi aider le passage de la perception des processus à la conception de la structure des modèles, notamment par test de leur sensibilité à la structure spatiale des réseaux, au mode de représentation du réseau (pas de discrétisation), au plan d'échantillonnage sur le réseau (taux de sondage, configuration spatiale) donc aux incertitudes de spatialisation, et ce pour différentes échelles. Ces simulations, au

même titre que d'autres méthodes de Monte-Carlo pourraient également être utilisées dans le cadre d'apprentissage de modèles, i.e. de calage et d'identification de paramètres de modèles, dans une approche de type GLUE (Beven et Binley, 1992), mais "spatiale".

*Utilisation hydrologique à plus long terme des méthodes développées*

Dans un cadre d'utilisation plus prospectif, si les déterminants des actes techniques sur ces aménagements étaient connus, l'évolution de la structure et des paramètres des modèles de variabilités de propriétés le long du réseau serait possible. L'intégration des réseaux simulés dans des modèles de changement, de prédiction hydrologique liée à des scénarios d'aménagement du bassin versant extrapolés serait alors envisageable.

Enfin, l'utilisation à terme de ces méthodes de spatialisation couplées à un modèle hydrologique distribué pour le diagnostic des impacts liés aux aménagements existants sur un bassin versant non "jaugé", dans une démarche d'ingénierie, est certainement le cadre d'utilisation le plus délicat et le plus lointain. Si cette utilisation en ingénierie hydrologique d'approches distribuées est encore assez peu d'actualité, on peut penser que ce virage "distribué" sera pris lorsque nous serons capables de représenter toutes les hétérogénéités spatiales influentes sur le fonctionnement du bassin versant.

Ce travail de recherche a contribué à cette phase de représentation des hétérogénéités spatiales internes d'un bassin versant.

La voie de généralisation de ce travail est désormais du côté de l'application hydrologique ou d'autres disciplines utilisatrices d'hétérogénéités spatiales de réseaux hydrographiques.



# Bibliographie

- AHERN, J. (2006). *From Landscape Research to landscape Planning : Aspects of Integration, Education and Application*, chapitre Theories, methods and strategies for sustainable landscape planning, pages 119–131. New-York : Springer-Verlag.
- AMBROISE, B. (1999). *La dynamique du cycle de l'eau*. H.G.A Bucarest.
- ANTOINE, J., BARACHE, D., CESAR, R. et da FONTOURA COSTA, L. (1997). Shape characterization with the wavelet transform. *Signal processing*, 62:265–290.
- BAILLY, J. (2002). Modelisation de processus à supports arborescents et orientés. Mémoire de D.E.A., Université de Montpellier II.
- BAILLY, J., LAGACHERIE, P., MILLIER, C., PUECH, C. et KOSUTH, P. (2006a). Comparison of artificial drainage network detection rates in cultivated landscapes with lidar and high resolution cir images. In 2006, E. G. U., éditeur : *Geophysical Research Abstracts 2006*, volume 8 de 1607-7962/gra/EGU06-A-08899.
- BAILLY, J., LAGACHERIE, P., MILLIER, C., PUECH, C. et KOSUTH, P. (in press). Agrarian landscapes linear features detection from lidar elevation profiles : application to artificial drainage network detection. *International Journal of Remote Sensing*.
- BAILLY, J., MONESTIEZ, P. et LAGACHERIE, P. (2006b). Modeling spatial variability along drainage networks with geostatistics. *Mathematical geology*, 38(5):515–539.
- BALTSAVIAS, E. (1999). Airborne laser scanning : existing systems and firms and other resources. *ISPRS Journal fo Photogrammetry and Remote Sensing*, 54:164–198.
- BARLOW, K., NASH, D., TURRAL, H. et GRAYSON, R. (2003). Phosphorus uptake and release in surface drain. *Agricultural water managment*, 63(2):109–123.
- BEASLEY, D. et HUGGINS, L. (1981). Answers (areal nonpoint source watershed environment response simulation). user's manual. report no. epa-905/9-82-001. Rapport technique, North Carolina State University.
- BEAULAND, A. (2003). Elaboration de modèles numériques de surface et de terrain en zone rurale fortement aménagée à partir de données lidar. Mémoire de D.E.A., Université de Paris VI.
- BENICHO, P. et LEBRETON, O. (1987). Prise en compte de la topographie pour la cartographie des champs pluviométriques statistiques. *Météorologie*, 19:1–28.

- BERGER, J. (2000). Les fossés bordiers historiques et l'histoire agraire rhodanienne. *Etudes rurales*, 153.
- BEVEN, K. (2001). How far can we go in distributed hydrological modelling? *Hydrology and Earth System Science*, 5:1–12.
- BEVEN, K. et BINLEY, A. (1992). The future of distributed models : Model calibration and uncertainty prediction. *Hydrological processes*, 6:279–298.
- BEVEN, K. et KIRKBY, M. (1979). A physically-based variable contributing area model of basin hydrology. *Hydrol. Sci. Bull.*, 24:27–53.
- BIARNES, A., RIO, P. et HOICHEUX, A. (2004). Analyzing the determinants of spatial distribution of weed control practices in a languedoc vineyard catchment. *Agro-nomie*, 24:187–196.
- BOUCHEREAU, E. B. (1997). *Analyse d'image par transformée en ondelette. Application aux images sismiques*. Thèse de doctorat, Université Joseph Fourier Grenoble.
- BRAUD, I., VARADO, N., GALLE, S., HAVERKAMP, R., LEBEL, T. et REGGIANI, P. (2001). Modélisation hydrologique du bassin versant de l'ouémé à l'aide du modèle power. In *Atelier sur le Couplage des modèles atmosphériques et hydrologiques*, Toulouse, France.
- BREIMAN, L. (1996a). Bagging predictors. *Machine learning*, 26(2):123–140.
- BREIMAN, L. (1996b). Random forest. *Machine learning*, 45:5–32.
- BREIMAN, L., FRIEDMAN, J., OLSHEN, R. et STONE, C. (1984). *Classification And Regression Tree*. London : Chapman and Hall.
- BRETAR, F. (2006). *Couplage de Données Laser Aéroporté et Photogrammétriques pour l'Analyse de Scènes Tridimensionnelles*. Thèse de doctorat, Ecole Nationale Supérieure des Télécommunications, Paris.
- BRODER, A. (1989). Generating random spanning trees. In *Proc. 30th Annual IEEE Symp. Foundations of Computer Science, IEEE, New York.*, pages 442–447.
- BRUCE, A. et GAO, H. (1996). *Applied wavelet analysis with S-plus*. New York : Springer-Verlag.
- BUREL, F. et BAUDRY, J. (2003). *Landscape ecology : concepts, methods, and applications*. Enfield : N.H. Science.
- CARLUER, N. et 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. *Journal of Hydrology*, 285(1):76–95.
- CARPENTIER, A., PAILLISSON, J., MARIONA, L., FEUNTEUN, E., BAISEZ, A. et RIGAUD, C. (2003). Trends of a bitterling (*rhodeus sericeus*) population in a man-made ditch network. *C. R. Biologies*, 326:166–173.
- CHAHINIAN, N. (2004). *Paramétrisation multi-critère et multi-échelle d'un modèle hydrologique spatialisé de crue en milieu agricole*. Thèse de doctorat, Université de Montpellier II.

- CHARLEUX-DEMARGNE, J. (2001). *Qualité des modèles numériques de terrain pour l'hydrologie. Application à la caractérisation du régime de crues des bassins versants*. Thèse de doctorat, Université de Marne-La-Vallée.
- CHENG, Z., ANDERSON, S., GANTZER, C. et CHU, Y. (2001). Fuzzy logic for predicting soil hydraulic conductivity using ct images. *Intelligent Engineering Systems Through Artificial Neural Networks Journal*, 11:307–312.
- CHILES, J. et DELFINER, P. (1999). *Geostatistics : modeling spatial uncertainty*. New York : John Wiley and Sons.
- COIFMAN, R. et DONOHO, D. (1995). *Wavelets and Statistics*, chapitre Translation invariant de-noising, pages 125–150. New York : Springer-Verlag.
- COLIN, F. (2000). *Approche spatiale de la pollution chronique des eaux de surface par les produits phytosanitaires. Cas du bassin du Sousson (Gers, France)*. Thèse de doctorat, ENGREF.
- COMORETTO, L. (2003). Analyse des chemins de l'eau d'un petit bassin rural sur mnt thr. Mémoire de D.E.A., Université de Montpellier II.
- CORTIJO, F. et de la BLANCA, N. P. (1997). A comparative study of some non-parametric spectral classifiers. applications to problems with high overlapping training sets. *International Journal of Remote Sensing*, 18(6):1259–1275.
- CRESSIE, N. (1993). *Statistics for spatial data*. John Wiley and Sons, Inc., New York.
- CRESSIE, N., FREY, J., HARCH, B. et SMITH, M. (2006). Spatial prediction on a river network. *Journal of Agricultural, Biological and Environmental Statistics*, 11(2):127–150.
- DAGES, C. (2006). *Analyse et modélisation de l'influence de réseaux de fossés sur les échanges surfacesouterrain en bassin versant méditerranéen*. Thèse de doctorat, Université de Montpellier II.
- DEFFONTAINES, B. (1990). *Développement d'une méthodologie morphonéotectonique - Analyse des surfaces enveloppes du réseau hydrographique et des MNT*. Thèse de doctorat, Université Paris VI.
- DELENNE, C. (2006). *Extraction et caractérisation des vignes à partir de données de télédétection à très haute résolution spatiale - Application en Languedoc-Roussillon pour la constitution de bases de données géographiques*. Thèse de doctorat, ENGREF.
- DIOT, O. (2000). Etude de la variabilité spatiale des réseaux d'écoulement anthropiques en milieu méditerranéen viticole : exemple de la basse vallée de la peyne. Mémoire de D.E.A., ULP Strasbourg.
- DODDS, P. et ROTHMAN, D. (1999). Unified view of scaling laws for river networks. *Phys. Rev. E*, 59(5):4865–4877.
- DOKOUMETZIDIS, A. et MACHERAS, P. (2003). A model for transport and dispersion in the circulatory system based on the vascular fractal tree. *Ann Biomed Eng.*, 31(3):284–293.



- DURUPT, M. et TAILLANDIER, T. (2006). Reconstruction automatique de bâtiments à partir d'un mne et de limites cadastrales : une approche opérationnelle. *In RFIA : Reconnaissance des Formes et Intelligence Artificielle, Tours, France.*
- EMERY, X. (2002). Conditional simulation of nongaussian random functions. *Mathematical geology*, 34(1):79–100.
- FAUCON, E. (1994). *Le parler de l'Hérault*. Lacour.
- FORTIN, J., TURCOTTE, R., MASSICOTTE, S. et FITZBACK, R. M. J. (2001). Distributed watershed model compatible with remote sensing and gis data. 1 : Description of the model. *Journal of the Hydrologic Division, ASCE*, 6(2):91–99.
- GALEA, G. et RAMEZ, P. (1995). *Maîtrise du ruissellement et de l'érosion en vignoble de coteaux*. Cemagref Editions.
- GINESTE, P., PUECH, C. et MÉROT, P. (1998). Radar remote sensing of the source areas from the coet-dan catchment. *Hydrological Processes*, 12:267–284.
- GIRARD, G., LEDOUX, E. et VILLENEUVE, J. (1981). Le modèle couplé, simulation conjointe des écoulements de surface et des écoulements souterrains sur un système hydrologique. *Cahiers ORSTOM, série Hydrologie*, XVIII(4):195–280.
- GOOVAERTS, P. (2001). Geostatistical modelling of uncertainty in soil science. *Geoderma*, 103:3–26.
- GOREAUD, F. (2000). *Apports de l'analyse de la structure spatiale en forêt tempérée à l'étude et la modélisation des peuplements complexes*. Thèse de doctorat, ENGREF.
- GRAYSON, R. et BLÖSCHL, G. (2000). *Spatial patterns in catchment hydrology : observations and modelling*, chapitre 3, pages 51–81. Cambridge : Cambridge University Press.
- GRAYSON, R., BLÖSCHL, G., WESTERN, A. et MCMAHON, T. (2002). Advances in the use of observed spatial patterns of catchment hydrological response. *Advances in Water Resources*, 25(8-12):1313–1314.
- GRAYSON, R., MOORE, I. et MAHON, T. M. (1992). Physically based hydrological modelling. 1. a terrain based model for investigative purposes. *Water Resources Research*, 28(10):2639–2658.
- GUÉDON, Y., BARTHÉLÉMY, D., CARAGLIO, Y. et COSTES, E. (2001). Pattern analysis in branching and axillary flowering sequences. *Journal of Theoretical Biology*, 212:481–520.
- GUIX, N. (2005). *Spatialisation du bilan hydrique de la vigne à l'échelle de la micro-région*. Thèse de doctorat, Université de Montpellier II.
- GUYON, X. (1992). *Champs aléatoires sur un réseau : modélisations, statistiques et applications*. Paris : Masson.
- HAAR, A. (1910). Zur theorie der orthogonalen funktionen-systeme. *Math. Ann.*, 69:331–371.

- HORNBERGER, G. et BOYER, E. (1995). Recent advances in watershed modelling : Us national report to the iugg 1991–1994. *Rev Geophys*, pages 949–957.
- HOWARD, A. (1967). Drainage analysis in geology interpretation : a summation. *Bull. Am. Assoc. Petrol. Geol.*, 51:2246–2259.
- ISTANBULLUOGLU, E., TARBOTON, D. et PACK, R. (2002). A probabilistic approach for channel initiation. *Water Resources Research*, 38(12):1–13.
- JANEY, N. (1992). *Modélisation et synthèse d'image d'arbres et de bassins fluviaux associant méthodes combinatoires et plongement automatique d'arbres et cartes planaires*. Thèse de doctorat, Université de Franche-Comté.
- JAVELLE, P. (2001). *Caractérisation du régime des crues : le modèle débit-durée-fréquence convergent. Approche locale et régionale*. Thèse de doctorat, INP Grenoble.
- JUTZI, B. (2005). Sub-pixel edge localization based on laser waveform analysis. In ENSCHEDE, T. n., éditeur : *IAPRS ISPRS WG III/3 Laserscanner 2005*, pages 109–114.
- KAO, C., VERNET, G., FILLEUL, J. L., NÉDÉLEC, Y., CARLUER, N. et GOUY, V. (2002). Élaboration d'une méthode de typologie des fossés d'assainissement agricole et de leur comportement potentiel vis-à-vis des produits phytosanitaires. *In-génieries*, 29:49–65.
- KLEMES, V. (1986). Dilettantism in hydrology : transition or destiny. *Water Resources Research*, 22:177–188.
- LACOSTE, C. (2004). *Extraction de réseaux linéiques à partir d'images satellitaires et aériennes par processus ponctuels marqués*. Thèse de doctorat, Université de Nice - Sophia Antipolis.
- LAGACHERIE, P., C.COLLIN-BELLIER et GOMA-FORTIN, N. (2001). Evaluation et analyse de la variabilité spatiale de la mortalité des ceps dans un vignoble languedocien à partir de photographies aériennes à haute résolution. *J. Int. Sci. Vigne Vin*, 35:141–148.
- LAGACHERIE, P., DIOT, O., DOMANGE, N., GOUY, V., FLOURE, C., KAO, C., MOUSSA, R., ROBBEZ-MASSON, J. et SZLEPER, V. (2004). 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.
- LANTUEJOUL, C. (2002). *Geostatistical Simulation : Models and Algorithms*. Springer Verlag.
- LECLERC, S. (2004). *Revisiter la notion de scénario hydrologique de référence pour la caractérisation des inondations*. Thèse de doctorat, Université Joseph Fourier de Grenoble.
- LECOEUR, C., BAUDRY, J., BUREL, F. et THENAIL, C. (2002). Why and how we should study field boundary biodiversity in an agrarian landscape context. *Agriculture, Ecosystems and Environment*, 89(1–2):23–40.

- LEFORNER, S. (2001). *Modélisation hydrologique des échanges surface-souterrain avec réseau de fossés en milieu agricole méditerranéen*. Thèse de doctorat, Université de Montpellier II.
- LEGENDRE, P. et LEGENDRE, L. (1998). *Numerical ecology : 2nd English edition*. Amsterdam : Elsevier Science BV.
- LEOPOLD, L. et MADDOCK, T. (1953). The hydraulic geometry of stream channels and some physiographic implications. *U.S. geol Surv. Prof. Pap.*, 252:9–16.
- LEPAPE, S. (1998). Analyse et quantification du réseau hydrographique. Mémoire de D.E.A., Ecole nationale des ingénieurs topographes.
- LESAFFRE, B. (1988). *Fonctionnement hydrologique et hydraulique du drainage souterrain des sols temporairement engorgés : Débits de pointe et modèle SIDRA*. Thèse de doctorat, Université de Paris VI.
- LOUCHART, X., VOLTZ, M., ANDRIEUX, P. et MOUSSA, R. (2001). Herbicides runoff at field and watershed scales in a mediterranean vineyard area. *Journal of Environmental Quality*, 30(3):982–991.
- LUDWIG, B., BOIFFIN, J., CHADOEUF, J. et AUZET, A. (1995). Hydrological structure and erosion damage caused by concentrated flow in cultivated catchments. *Catena*, 25:227–252.
- MADRIGAL, I., BENOIT, P., E. BARRIUSO, V. E., SOUILLER, C., REAL, B. et DUTERTRE, A. (2002). Capacités de stockage et d'épuration des sols de différents dispositifs enherbés vis-à-vis des produits phytosanitaires. deuxième partie : Propriétés de rétention de deux herbicides, l'isoproturon et le diflufenicanil dans différents sols de bandes enherbées. *Etude et gestion des sols*, 9(4):287–302.
- MALLAT, S. (1989). A theory for multi-resolution signal decomposition : the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–693.
- MANLY, B. (1997). *Randomization, bootstrap and Monte-Carlo methods in biology, 2nd edition*. Chapman and Hall, London.
- MANNA, S. (1998). Branched tree structures : From polymers to river networks. *Physica A*, 254:190–197.
- MARGOUM, C., GOUY, V., WILLIAMS, R. et SMITH, J. (2001). Le rôle des fossés agricoles dans la dissipation des produits phytosanitaires. *Ingénieries*, pages 56–65.
- MARKS, K. et BATES, K. (2000). Integration of high-resolution topographic data with floodplain models. *Hydrologic processes*, 14:2109–2122.
- MAROFI, S. (1999). *Rôle des échanges nappes-fossés dans le fonctionnement hydrologique d'un bassin versant en milieu méditerranéen cultivé*. Thèse de doctorat, ENSA Montpellier.
- MAROFI, S., BOUZIGUES, R., MOUSSA, R. et VOLTZ, M. (2002a). Dynamic of water exchanges between ditches and shallow groundwater in a mediterranean catchment.

- MAROFI, S., VOLTZ, M. et MOUSSA, R. (2002b). Role of ditches on water exchanges between two groundwaters in an agricultural catchment. *In 17th WCSS, Thaïland.*
- MASON, D., ANDERSON, G. et BRADBURY, R. (2003). Measurement of habitat predictor variables for organism-habitat models using remote sensing and image segmentation. *International Journal of Remote Sensing*, 24(12):2515–2532.
- MENDES-FRANCE, M. (1981). De l'arbre de leonardo da vinci à la théorie de la dimension. *Revue du Palais de la Découverte*, 10:52–60.
- MOLZ, F., RAJARAM, H. et LU, S. (2004). Stochastic fractal-based models of heterogeneity in subsurface hydrology : Origins, applications, limitations, and future research questions. *Stochastic Environmental Research and Risk Assessment (SERRA)*, 18(4):278–279.
- MONESTIEZ, P., BAILLY, J., LAGACHERIE, P. et VOLTZ, M. (2005). Geostatistical modelling of spatial processes on directed trees : Application to fluvisol extent. *Geoderma*, 128, 3-4:179–191.
- MONESTIEZ, P., HABIB, R. et AUDERGON, J. (1989). Estimation de la covariance et du variogramme pour une fonction aleatoire a support arborescent : application à l'étude des arbres fruitiers. *Geostatistics*, 1:39–56.
- MOUSSA, R. (submitted). Effect of channel network topology, basin segmentation and rainfall spatial distribution on the giuh transfer function. *Journal of Hydrology*.
- MOUSSA, R. et BOCQUILLON, G. (1996). Criteria for the choice of flood-routing methods in natural channels. *Hydrological Processes*, 186(1-4):1–30.
- MOUSSA, R., VOLTZ, M. et 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.
- NÉDELEC, Y. (2005). *Interactions entre les fonctionnements hydrauliques des réseaux enterrés et à surface libre, dans un bassin versant drainé, lors du transfert des crues.* Thèse de doctorat, ENGREF.
- NIEMANN, J., BRAS, R. L., VENEZIANO, D. et RINALDO, A. (2001). Impacts of surface elevation on the growth and scaling properties of simulated river networks. *Geomorphology*, 40(1-2):37–55.
- OBERLIN, G. et POULARD, C. (2002). Programme risque inondation. dies irae : Diagnostic des effets d'une stratégie intégrée de ralentissement dynamique des écoulements sur le régime des crues. Rapport technique, Cemagref Lyon HHLY. 163p.
- OLLIER, C. et POIRÉE, M. (1981). *Assainissement agricole.* Paris : Eyrolles.
- OLLIER, S., CHESSEL, D., COUTERON, P., PÉLISSIER, R. et THIOULOUSE, J. (2003). Comparing and classifying one-dimensional spatial patterns : an application to laser altimeter profile. *Remote sensing of environment*, 85:453–462.
- PAOLI, J. (2005). *Fusion de données spatialisées. Application à la Viticulture de Précision.* Thèse de doctorat, Université de Montpellier II.

- PECKHAM, S. D. (1995). New results for self-similar tree with application to river networks. *Water Resour. Res.*, 31(4):1023–1029.
- PENAUD, J. (1988). The ramification matrix of random binary trees. rapport de recherche, labri n. 8832. Rapport technique, Département d'Informatique, Université de Bordeaux I.
- PEREIRA, L. G. et WICHERSON, R. (1999). Suitability of laser data for driving geographical information. a case study in the context of management of fluvial zone. *ISPRS Journal of Photogrammetry and Remote Sensing*, 54:105–114.
- PITA, R., MIRA, A. et BEJA, P. (2006). Conserving the cabrera vole, *Microtus cabrerae*, in intensively used mediterranean landscapes. *Agriculture, Ecosystems and Environment*, 115(3):1–5.
- RACLOT, D. et COLIN, F. (2005). Updating land cover classification using a rule-based decision system. *International Journal of Remote Sensing*, 26(7):1309–1321 ;.
- RACLOT, D. et PUECH, C. (2002). What does ai contribute to hydrology? exemple on aerial photos and flood levels. *Applied Artificial Intelligence*, 17(1):71–86.
- RIPLEY, B. (1987). *Stochastic Simulation*. John Wiley and Sons, Inc., New York.
- RITCHIE, J., GRISSINGER, E., MURPHEY, J. et GARBRECHT, J. (1994). Measuring channel and gully cross-sections with an airborne laser altimeter. *Hydrological Processes*, 7:237–244.
- ROUQUETTE, J. et THOMPSON, D. J. (2005). Habitat associations of the endangered damselfly, *Coenagrion mercuriale*, in a water meadow ditch system in southern england. *Biological Conservation*, 123:225–235.
- ROUVIÈRE, L. (2005). *Estimation de densités en dimensions élevées et classifications de courbes*. Thèse de doctorat, Université de Montpellier II.
- SAIDOUNI, D. (2003). Qualité et classification de données lidar multipulsés pour l'extraction de réseaux d'écoulement en bassin rural aménagé. Mémoire de D.E.A., Université Louis Pasteur de Strasbourg.
- SALAS, E., OZIER-LAFONTAINE, H. et NYGREN, P. (2004). A fractal root model applied for estimating the root biomass and architecture in two tropical legume tree species. *Ann. For. Sci.*, 61:337–345.
- SCHEIDEGGER, A. E. (1991). *Theoretical Geomorphology*. Springer-Verlag, New York.
- SNELL, J. et SIVAPALAN, M. (1994). On geomorphological dispersion in natural catchments and the geomorphological unit hydrograph. *Water Resources Research*, 4:2311–2323.
- SOILLE, P. (1999). *Morphological Image Analysis : Principles and Applications*. New York : Springer-Verlag.
- STEINITZ, C. (1990). A framework for theory applicable to the education of landscape architects (and other environmental design professionals). *Landscape Journal*, 9(2): 136–143.

- STOICA, R. (2001). *Processus ponctuels pour l'extraction de réseaux linéiques dans les images satellitaires et aériennes*. Thèse de doctorat, University de Nice-Sophia Antipolis.
- STRAHLER, A. N. (1952). Hypsometric (area-altitude) analysis of erosional topology. *Bull. Geol. Soc. Am*, 63:1117–1142.
- STRUZIK, Z. (2000). Wavelet methods in financial time-series processing. *Physica A*, 296:307–319.
- SZLEPER, V. (2002). Utilisation d'une approche par indicateur pour caractériser les réseaux d'écoulement anthropiques vis à vis des pollutions par pesticides. Mémoire de D.E.A., ENGEES.
- TAKAYASU, H. (1990). *Fractals in the Physical Sciences*. Manchester University Press, Manchester.
- TARBOTON, D. (1997). A new method for the determination of flow directions and upslope areas in grid digital elevation models. *Water resources research*, 33(2):309–319.
- TEMPLE-BOYER, E. (2000). Les talus, éléments linéaires du paysage méditerranéen (bassin versant de roujan, hérault). Mémoire de D.E.A., UFR de Géographie, Université de Paris I.
- THORNTON, M., ATKINSON, P. et HOLLAND, D. (2006). Sub-pixel mapping of rural land cover objects from fine spatial resolution satellite sensor imagery using super-resolution swapping. *International Journal of Remote Sensing*, 27(3):473–491.
- TRIAS-SANZ, R. et DESEILLIGNY, M. P. (2004). A region-based method for graph to image registration with an application to cadastre data. *In ICIP*, pages 1703–1706.
- TRIPATHI, N., GOKHALE, K. et SIDDIQUI, M. (2000). Directional morphological image transforms for lineament extraction from remotely sensed images. *International Journal of Remote Sensing*, 21(17):3281–3292.
- TURCOTTE, R., FORTIN, J., ROUSSEAU, A., MASSICOTTE, J. et VILLENEUVE, J. (2001). Determination of the drainage structure of a watershed using a digital elevation model and a digital river and lake network. *Journal of hydrology*, 240:225–242.
- VERHOEF, J., CRESSIE, N. et BARRY, R. (2004). Flexible spatial models for kriging and cokriging using moving average and the fast fourier transform (fft). *Journal of computational and graphical statistics*, 13:265–282.
- VERHOEF, J., PETERSON, E. et THEOBALD, D. (2006). Spatial statistical models that use flow and stream distance. *Environmental and Ecological Statistics*, 13:449–464.
- VERTESSY, R., HATTON, T., O'SHAUGHNESSY, P. et JAYASURIYA, M. (1993). Predicting water yield from a mountain ash forest catchment using a terrain analysis based catchment model. *Journal of Hydrology*, 150:665–700.
- WAASENAAR, T. (2001). *Soil surface feature detection in the French vine-cultivated Mediterranean region by remote sensing at very high spatial resolution*. Thèse de doctorat, ENSAM.

- WANG, L. et SASSEN, K. (2002). Cirrus cloud microphysical property retrieval using lidar and radar measurements : I. algorithm description and comparison with in situ data. *J. Appl. Meteor.*, 41:218–229.
- WATSON, A. M. et ORMEROD, S. (2004). The distribution of three uncommon freshwater gastropods in the drainage ditches of british grazing marshes. *Biological Conservation*, 118:455–466.
- WESTERN, A., BLÖSCHL, G. et GRAYSON, R. (2001). Towards capturing hydrologically significant connectivity in spatial patterns. *Water Resources*, 37(1):83–97.
- WESTERN, A. W., ZHOUA, S., GRAYSON, R., MCMAHON, T. A., BLOSCHL, G. et WILSON, D. (2004). Spatial correlation of soil moisture in small catchments and its relationship to dominant spatial hydrological processes. *Journal of Hydrology*, 286:113–134.
- WIECHERT, A., KATZENBEISSER, R. et KURZ, S. (2000). Precision lidar data and true ortho-images : a comparison of airborne laser scanning with terrestrial and photogrammetry. [www.toposys.com](http://www.toposys.com).

# Annexe A

## Annexes

### Sommaire

---

<b>A.1 Annexes du chapitre 3</b>	<b>169</b>
A.1.1 Directed path definition	169
A.1.2 Rivers ordering	169
A.1.3 Conditional independence for geostatistical models on directed tree	169
A.1.4 Simulation of random function on a directed tree	170
A.1.5 Illustration de la procédure de simulation conditionnelle aval-amont sur une rivière	171
<b>A.2 Annexes du chapitre 4</b>	<b>174</b>
A.2.1 Présentation du système et des données LiDAR	174
A.2.1.1 Principes physiques du LiDAR	174
A.2.1.2 Les LiDAR topographiques aéroportés	174
A.2.1.3 Les principes d'enregistrement d'échos laser	175
A.2.1.4 Les données LiDAR Toposys du système Falcon II	176
A.2.2 Transformée en ondelettes 1D : Quelques principes et exemples	179
A.2.2.1 Exemple d'une transformation en ondelettes discrète (D.W.T) pas à pas	179
A.2.2.2 Exemples virtuels de concavités et coefficients d'ondelettes de Haar associés	179
A.2.2.3 Comparaison de la transformée en ondelettes décimées (D.W.T) et continues (S.W.T)	181
A.2.3 A.C.P sur les coefficients d'ondelettes : interprétation et sélection des facteurs	184
A.2.3.1 Interprétation des composantes principales	184
A.2.3.2 Sélection des composantes discriminantes	185
A.2.4 Effet du couple d'ondelettes sur les résultats de classification des profils altimétriques	186
A.2.5 Algorithme CART	187
A.2.5.1 Expansion de l'arbre sur l'échantillon d'apprentissage	187
A.2.5.2 Élagage de l'arbre sur un échantillon test	188



A.2.6	Comparaison de performances de classification de profils altimétriques LiDAR suivant les algorithmes CART et RANDOM FOREST . . . . .	189
A.2.6.1	Introduction . . . . .	189
A.2.6.2	Données . . . . .	189
A.2.6.3	Méthodes . . . . .	189
A.2.6.3.1	Arbre de classification CART sélectionné par rééchantillonnage aléatoire . . . . .	189
A.2.6.3.2	Random Forest . . . . .	189
A.2.6.4	Résultats . . . . .	190
A.2.6.4.1	Sur le site <b>B</b> . . . . .	190
A.2.6.4.2	Sur le site <b>C</b> . . . . .	191
A.2.6.5	Conclusion . . . . .	192
A.2.7	Détection de fossés par classification de profils radiométriques issus d'images optiques à très haute résolution spatiale . . . . .	194
A.2.7.1	Methods . . . . .	194
A.2.7.2	Images data . . . . .	194
A.2.7.2.1	CIR panchromatic image <b>I1</b> . . . . .	194
A.2.7.2.2	CIR panchromatic image <b>I2</b> . . . . .	194
A.2.7.3	Visual interpretation . . . . .	194
A.2.7.4	Automated classification . . . . .	195
A.2.7.5	Conclusion . . . . .	196
A.2.8	Classification binaire des arêtes du lattice cadastral à partir des profils LiDAR classés suivant des modèles "dissymétriques" . . . . .	197
A.2.8.1	Lissage des résultats de classification de profils EXT-ND . . . . .	197
A.2.8.2	Lissage des résultats de classification de profils EXT-D . . . . .	197
A.2.9	Loi de Tokunaga . . . . .	201
A.2.10	Exemples de lois géométriques de longueurs des branches . . . . .	201

---

## A.1 Annexes du chapitre 3

Les trois premières sections suivantes sont les annexes de l'article publié dans *Mathematical Geology* intitulé "Modeling spatial variability along drainage network with geostatistics" (Bailly *et al.*, 2006b).

### A.1.1 Directed path definition

A directed path  $L$  of size  $q$  is a sequence of  $q$  edges,  $L = \{u_1, u_2, u_3, \dots, u_q\}$  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  $L$ .

### A.1.2 Rivers ordering

The whole drainage network can be decomposed in a unique set of river  $A_o = \{a_1, a_2, \dots, a_{N_S}\}$ , 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  $l_{b_i r_G}$ .

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

Let  $A_o = a_{(1)}, a_{(2)}, \dots, a_{(N_S)}$  be an ordered sequence of  $N_S$  rivers (Fig. 3.2B) which is sequentially defined by

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

$A_o$  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.

### A.1.3 Conditional independence for geostatistical models on directed tree

Conditional independence between parallel rivers is defined by the equality of two conditional distributions :

$\forall n \in \mathbb{N}, \quad \forall s_1, \dots, s_n \in G_M, \quad \forall s' \in G_M,$  the set of  $s_i$ 's is arranged in three subsets reordering the indices so that  $s_1, \dots, s_\alpha$  belong to  $\Gamma_{s'}^{-1}$ ,  $s_{\alpha+1}, \dots, s_\beta$  belong to  $\Gamma_{s'}$  and  $s_{\beta+1}, \dots, s_n$  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  $s_\beta$ . Then :

$$\mathcal{L}\left(Z(s_1), \dots, Z(s_\alpha) | Z(s_{\beta+1}), \dots, Z(s_n), Z(\Gamma_{s'})\right) = \mathcal{L}\left(Z(s_1), \dots, Z(s_\alpha) | Z(\Gamma_{s'})\right) \quad (\text{A.1})$$

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 :

$$\begin{aligned} &\forall n \in \mathbb{N}, \quad \forall z_1, \dots, z_n \in \mathbb{R}^n, \\ &\forall L, \quad \forall s_1, \dots, s_n \in L \text{ and } \forall s'_1, \dots, s'_n \in L \text{ such as } s'_1 \in \Gamma_{s_1}, \dots, s'_n \in \Gamma_{s_n} \\ &\text{and such as } d(s_1, s'_1) = h, \dots, d(s_n, s'_n) = h \end{aligned}$$

$$P\left(Z(s_1) < z_1, \dots, Z(s_n) < z_n\right) = P\left(Z(s'_1) < z_1, \dots, Z(s'_n) < z_n\right) \quad (\text{A.2})$$

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

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

#### A.1.4 Simulation of random function on a directed tree

Let  $\mathcal{D} = \{s_1, \dots, s_\alpha, \dots, s_N\}$  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  $a_{(1)}, a_{(2)}, \dots, a_{(N_S)}$  of  $A_o$ , we denote  $\mathcal{D}$  using new subscripts

$$\mathcal{D} = \{s_{(1)}^1, \dots, s_{(n_1)}^1, s_{(1)}^2, \dots, s_{(n_2)}^2, \dots, s_{(1)}^{N_S}, \dots, s_{(n_{N_S})}^{N_S}\}$$

where the subset  $\{s_{(1)}^k, \dots, s_{(n_k)}^k\}$  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  $\{s_{(1)}^1, \dots, s_{(n_1)}^1\}$  belonging to  $a_{(1)}$ .

Let  $Y_{a_{(1)}}$  denote a random vector  $Y s_{(1)}^1, \dots, Y s_{(n_1)}^1$ . 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  $s_\alpha$  and  $s_{\alpha'} \in \{s_{(1)}^1, \dots, s_{(n_1)}^1\}$ .

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

$$Y_{a_{(1)}} = L_{(1)} \varepsilon_{(1)} \quad (\text{A.3})$$

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  $\Sigma_{(0)} = L_{(0)} \cdot {}^t L_{(0)}$ .

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 $A_o$

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

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

The first one  $D_1$  contains the points to simulate  $D_1 = \{s_{(1)}^k, \dots, s_{(n_k)}^k\}$ , the second  $D_2$  the points already simulated  $D_2 = \{s_{(1)}^1, \dots, s_{(n_1)}^1, \dots, s_{(1)}^{k-1}, \dots, s_{(n_{k-1})}^{k-1}\}$  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 (A.1), it follows for the following distributions :

$$\left( Y(s_\alpha), s_\alpha \in D_1 \mid Y(s_\nu), s_\nu \in D_2 \right) \sim \left( Y(s_\alpha), s_\alpha \in D_1 \mid Y(s_\beta), s_\beta \in D_3 \right)$$

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

$$\forall s_\alpha \in a_{(k)}, \quad Y(s_\alpha) = Y_{nc}(s_\alpha) + \left( Y^*(s_\alpha) - Y_{nc}^*(s_\alpha) \right) \quad (\text{A.4})$$

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  $(Y(s_1), \dots, Y(s_{n_k}))$  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.

### A.1.5 Illustration de la procédure de simulation conditionnelle aval-amont sur une rivière

On représente sur la figure A.1 la procédure de simulation détaillée dans la sous-section précédente : Cet exemple porte sur une rivière du réseau hydrographique artificiel de Roujan dont la topologie est représentée en haut de cette figure. Cette rivière, dite rivière numéro 30, apparaît en ligne rouge large sur cette figure. On prend ensuite comme points de données, 43 points situés en aval de la confluence située sur cette rivière. On suppose ces points de données comme des données déjà simulées dans une étape précédente, car situés en aval d'une confluence. Les trois sous-graphiques du bas de la figure A.1 représentent les calculs intermédiaires de la simulation conditionnelle de  $Z$  (largeur) en amont de ces points de données, i.e. en amont de la confluence. Dans un premier temps (en haut), on fait une simulation non conditionnelle (courbe noire) du résidu, notée  $Y_{nc}$  sur toute la rivière. Aux emplacements de points de données, ces valeurs de résidus simulés sont les points oranges. On réalise ensuite un krigeage simple (moyenne nulle) à partir de ces points oranges : on obtient la courbe rouge représentant  $Y^*_{nc}$ . On remarque qu'en extrapolation ce

résidu krigé tend vers l'espérance du champ, soit 0. La différence entre la simulation non conditionnelle et le krigeage simple, notée  $Y_{nc} - Y^*_{nc}$ , est ensuite représentée dans le graphique du milieu en gris. On réalise ensuite un krigeage simple à partir des vrais résidus calculés sur les points de données (courbe orange), noté  $Y^*$ . On somme ensuite  $Y^*$  et  $Y_{nc} - Y^*_{nc}$  pour obtenir  $Y = Y^* + (Y_{nc} - Y^*_{nc})$  représenté par la courbe verte : on obtient ainsi une simulation conditionnelle du résidu  $Y$  tout le long de la rivière 14 à partir des résidus situés en aval de la confluence. Ces résidus respectent la distribution multi-gaussienne du résidu, la structure de covariance exponentielle et les 43 points de données en aval. On voit bien sur cet exemple, qu'à partir d'une distance aux données vers l'amont égale à la portée du variogramme, le conditionnement disparaît (les courbes vertes et grises se confondent). Le dernier graphique en bas représente la simulation conditionnelle de  $Z$  sur toute la rivière 30 (par rapport aux 43 points de données sur  $Z$  en aval) en ajoutant le terme de dérive représentée par la courbe rouge. On perçoit sur ce graphique et pour une rivière, l'effet assez faible de la dérive et les discontinuités aux confluences induites par le modèle de dérive (légers décrochements de la courbe vers 880 m correspondant à la confluence sur la rivière 30).

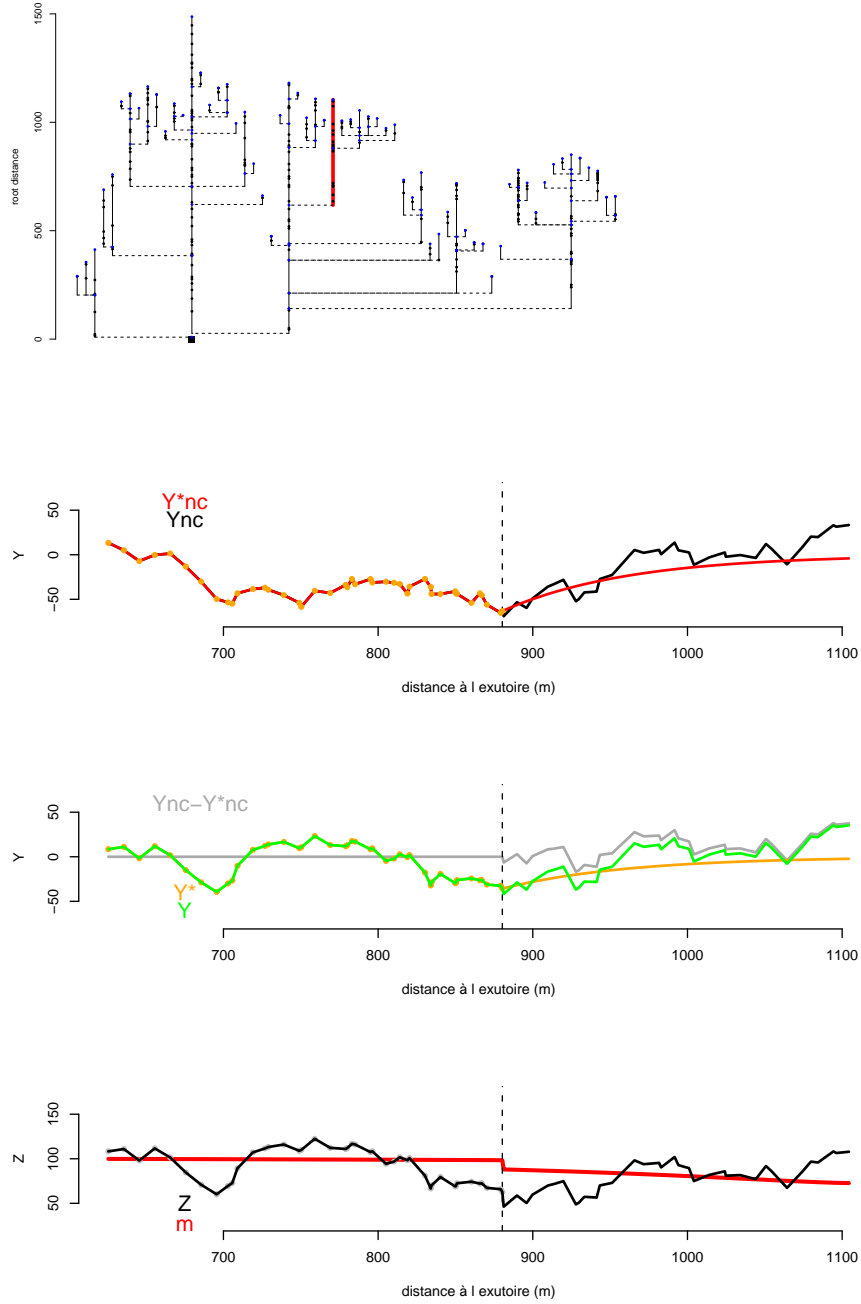


FIG. A.1 – Exemple de simulation conditionnelle aval-amont sur la rivière numéro 30 : 1-Haut : emplacement de la rivière 30 dans le réseau (en rouge) 2- Bas : Étapes successives de la simulation conditionnelle : les données de conditionnement sont celles situées en aval de la confluence (trait pointillé vertical).

## A.2 Annexes du chapitre 4

### A.2.1 Présentation du système et des données LiDAR

#### A.2.1.1 Principes physiques du LiDAR

On appelle LiDAR, pour *Light Detection and Ranging*, ou "radar à lumière", un système d'émission-réception d'une impulsion de lumière générée par un laser (Fig. A.2). Ces systèmes peuvent être classés en plusieurs catégories : LiDAR à intensité, généralement multispectral (fluorescence, raman... : cas des LiDAR atmosphériques), LiDAR Doppler (mesures de vents et vitesses de cibles) et LiDAR "Range Finder" ou LiDAR de "télémétrie". Ces derniers sont ceux utilisés pour effectuer des levés topographiques soit par des systèmes à balayage au sol à très grande échelle, soit depuis des plate-formes aéroportées ou depuis un satellite (ICESAT, CALYPSO). Le principe de la télémétrie LiDAR aéroportée est relativement simple. On transforme en distance une durée entre l'instant d'émission d'une impulsion lumineuse et l'instant de réception de la part réfléchi de cette impulsion par la cible. Cette transformation est simple lorsque la vitesse de la lumière est constante dans le milieu traversé par le faisceau (ce qui n'est pas le cas par exemple des LiDAR dit bathymétriques qui traverse une colonne d'air puis d'eau de célérités différentes).

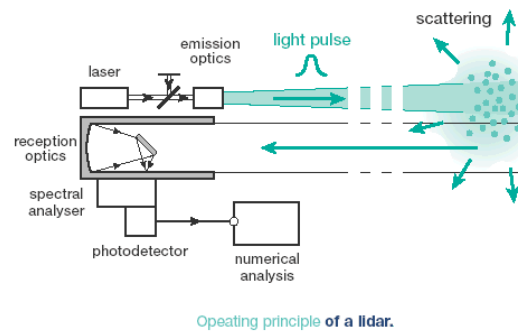


FIG. A.2 – Principe du LiDAR, source "[smc.cnes.fr/CALIPSO/GP.mission.htm](http://smc.cnes.fr/CALIPSO/GP.mission.htm)"

#### A.2.1.2 Les LiDAR topographiques aéroportés

Les systèmes LiDAR utilisés pour effectuer des relevés topographiques de surfaces émergées depuis une plate-forme aéroportée (avion, hélicoptère) ont des caractéristiques communes (Baltasvias, 1999), à savoir :

- L'émission des impulsions laser est effectuée généralement par des lasers à matrice cristalline Nd :Yag<sup>1</sup> à très haute fréquence d'émission (jusque 350 Khz) qui émettent dans le Proche-Infra-Rouge (1064 nm). Les impulsions laser sont généralement des ondes gaussiennes sur une échelle temps-puissance.
- Afin de couvrir une bande de terrain 'large' par axe de vol de l'avion et optimiser le temps de vol, un système de balayage latéral des faisceaux est couplé en sortie du laser. Ces systèmes sont soit mécaniques (miroirs oscillants, pivotants), soit "fixes" (réseau de fibres optiques)

<sup>1</sup>Excitation par pompage d'une matrice de grenat d'aluminium et d'yttrium par du néodyme

- La qualité de positionnement du point topographique, cible du faisceau laser, dépend bien sûr de la qualité de la mesure de distance entre l'avion (ou l'hélicoptère) et cette cible mais surtout de la qualité de la géométrie de la visée du laser à cet instant (lacet, roulis, tangage de l'avion, position du miroir. . .) et de la qualité du positionnement de la plate-forme aéroportée à cet instant. Ces variables sont déterminées par un positionnement D-GPS cinématique de haute précision qui impose de voler à proximité (moins de 30 km) d'une station de référence D-GPS, et par des centrales inertielles (accéléromètres) embarquées qui permettent de mesurer tous les effets de navigation (roulis, tangage, lacet. . .). Cependant, ces instruments étant à plus basses fréquences d'acquisition (1 à 100 Hz) que le laser, ils représentent souvent le point "faible" de la qualité de positionnement topographique des levés topographiques LiDAR.

Ces systèmes se distinguent cependant par plusieurs paramètres qui jouent sur la qualité et la résolution de la mesure topographique :

- la puissance de l'impulsion laser. Pour une même altitude de vol, plus l'énergie de l'impulsion est importante, plus le pouvoir de pénétration du faisceau laser est important au travers des objets de sursol (végétation) rencontrés à la surface.
- la fréquence d'émission du laser. Pour une vitesse d'avion donnée, plus elle est importante, plus la densité de points échantillonnés au sol est grande.
- l'angle de divergence du faisceau, appelé également angle d'ouverture du faisceau laser. Plus il est faible et plus, pour une altitude de vol donnée, l'emprunte du faisceau au sol sera fine. Actuellement, ces empruntes vont de 2 m à 5 cm (système Flymap-Fugro).
- la fréquence de numérisation sur le récepteur,
- le mode d'enregistrement du retour de l'onde laser : mono-écho (le premier écho laser significatif), multi-échos (souvent deux : premier-deuxième), retour d'onde complète.

Ce dernier point permet d'effectuer une taxonomie des systèmes LiDAR actuels utilisés pour l'étude et la topographie des surfaces continentales.

### A.2.1.3 Les principes d'enregistrement d'échos laser

Chaque impulsion laser émise, en fonction de la transmittance des milieux traversés, en fonction des propriétés de réflectance et d'absorption des cibles rencontrées par le faisceau renvoie vers le capteur une onde modifiée appelée retour d'onde complète (Fig. A.3-droite), exprimée sur une échelle temps-puissance. On parle également de temps-intensité. Typiquement, on obtient un pic de puissance (intensité) par "objet" rencontré verticalement par le faisceau (branche d'arbre), tant qu'il n'est pas éteint (Fig. A.3).

Pour un problème de compromis entre densité d'échantillonnage au sol et stockage de l'information d'intensité, cette forme d'onde, lorsqu'elle n'est pas complètement enregistrée (cas de certains capteurs bathymétriques et capteurs dits "full wave form") dans le cas des applications purement topographiques, est analysée suivant trois grands modes qui "seuillent" ces ondes complètes :

1. Mono-écho : seul le premier pic de puissance est enregistré. Il correspond aux premières cibles du faisceau, i.e. le toit de la végétation. Ce mode permet de travailler à très haute fréquence avec des faisceaux précis, à faible divergence (système Flymap-Fugro)



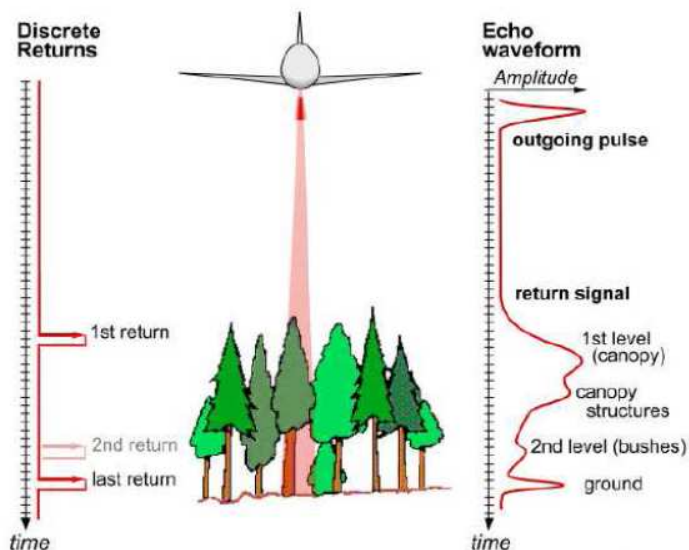


FIG. A.3 – Exemple de retour d’onde complète (Droite) et illustration d’un mode multi-échos (gauche).

2. Bi-échos : on enregistre le premier et le deuxième (si différent du premier) pic de puissance.
3. Multi-échos : on choisit d’enregistrer un nombre de pics significatifs plus grand que deux.

Pour tout ces systèmes, l’onde de retour est seillée suivant différentes méthodes (maximas locaux, centre de gravité au delà d’une certaine puissance. . .) (Bretar, 2006).

Suivant la dimension de l’empreinte au sol du faisceau et suivant la géométrie de la cible, les échos peuvent représentés également des différences de hauteurs "latérales" de la cible, comme le montrent les exemples de la figure A.4 (Jutzi, 2005).

#### A.2.1.4 Les données LiDAR Toposys du système Falcon II

Des données LiDAR ont été acquises sur le bassin du Bourdic le 26 juin 2002 avec le système FALCON-II de TOPOSYS depuis un hélicoptère. Ce système est un LiDAR topographique PIR bi-échos (First-Last Pulse), avec un système de balayage par fibres optiques. Les principales caractéristiques sont indiquées sur le tableau A.1.

L’utilisation d’un hélicoptère permet une grande densité de points : la densité spatiale de points de mesures est d’environ un point tous les 10 cm dans le sens de l’axe de vol et de un point tous les 2 m perpendiculairement à l’axe de vol. Du fait du système de fibres optiques utilisé pour le balayage, l’échantillonnage spatial est composé de courbes parallèles à l’axe de vol qui oscillent (mode 'swing') pour former des sinusoides parallèles (Fig. A.5).

Sur ces données, le dernier écho (points noirs sur l’exemple en figure A.6) correspond, soit :

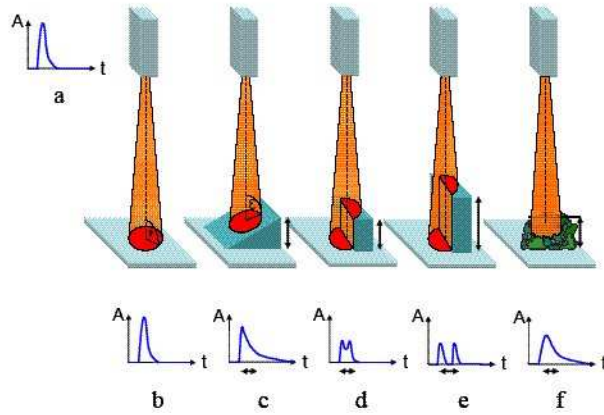


FIG. A.4 – Effet de la surface sur le train d’ondes ; a) Train d’ondes transmis b) Surface plane c) Surface avec une pente d) Deux surfaces séparées par une légère élévation e) Deux surfaces séparées par une grande élévation f) Distribution aléatoire de petits objets, d’après Jutzi (2005).

Altitude du vecteur (hélicoptère)	900 m
Fauchée par axe de vol	250 m
Durée d’une impulsion	4 ns
Fréquence d’acquisition	83 kHz
Longueur d’onde du laser (PIR)	1.55 $\mu\text{m}$
Dimensions du point laser au sol	$\simeq 0.6 \times 0.6$ m

TAB. A.1 – Caractéristiques de la campagne LiDAR Falcon II-Toposys du 26 juin 2002 réalisé sur le bassin du Bourdic

- au toit de la végétation lorsque celle-ci est si dense qu’elle ne laisse pas pénétrer le faisceau laser, dans ce cas les deux échos sont confondus (il n’y a pas de dernier écho) : toit des houppiers de vigne ;
- la végétation est variable sur la dimension de l’empreinte laser au sol (disque de 0.6 m de diamètre environ dans notre cas) et l’écho intègre ces différences de végétation (exemple des points en bord de houppier de vigne) ;
- soit au terrain (inter-rang de vigne).

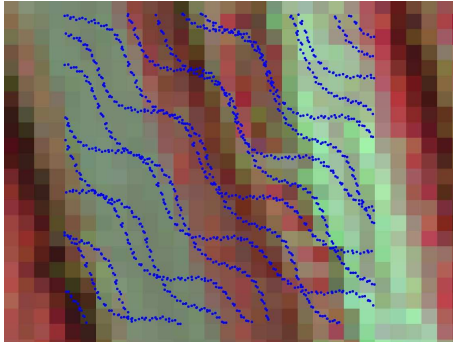


FIG. A.5 – Forme de l'échantillonnage des points LiDAR sur image I.R.C. à 0.5 m de résolution. Les sinusoides sont doubles car nous sommes sur une zone de recouvrement de 2 axes de vol.

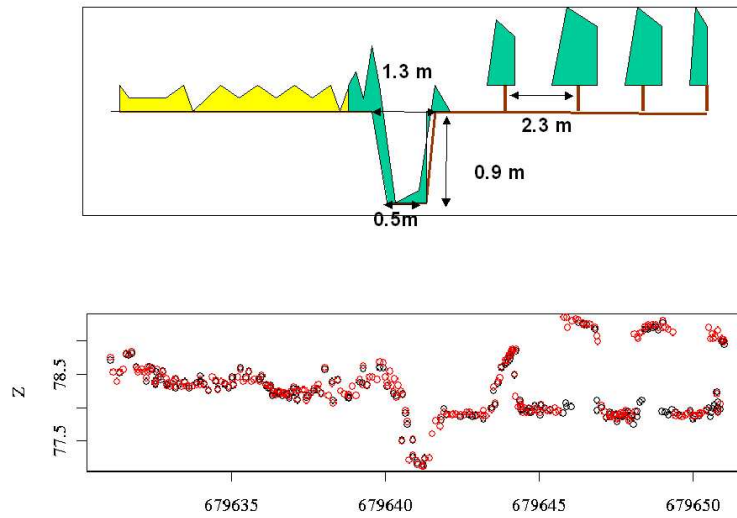


FIG. A.6 – Exemple de transect des points LiDAR premier écho (rouge) et dernier écho (noir) réalisés à partir des données Toposys sur le bassin du Bourdic.

## A.2.2 Transformée en ondelettes 1D : Quelques principes et exemples

### A.2.2.1 Exemple d'une transformation en ondelettes discrète (D.W.T) pas à pas

Les fonctions  $\phi_{J,k}(t)$  et  $\psi_{j,k}(t)$  de l'équation 4.1 (p. 102) sont construites suivant des résolutions dyadiques<sup>2</sup> comme expliqué ci-après, par translation et dilatation :

$$\phi_{J,k}(t) = 2^{-\frac{J}{2}} \phi\left(\frac{t - 2^J k}{2^J}\right) \quad (\text{A.5})$$

et

$$\psi_{j,k}(t) = 2^{-\frac{j}{2}} \phi\left(\frac{t - 2^j k}{2^j}\right) \quad (\text{A.6})$$

Les coefficients  $S_{J,k}, d_{J,k}, \dots, d_{1,k}$  désignent les coefficients de transformée en ondelettes.  $S_{J,k}$  correspondent aux coefficients de "tendance" (ou père) et  $d_{j,k}$  correspondent aux coefficients de détails (ou mère). Dans un cas de transformée continue, les coefficients  $d_{j,k}$  sont approximativement les intégrales :

$$d_{j,k} \approx \int \psi_{j,k}(t) z(t) dt \text{ for } j = 1, \dots, J \quad (\text{A.7})$$

La valeur absolue de chaque coefficient mesure la contribution de la fonction ondelette correspondante dans l'approximation globale (Eq. 4.1) (Bruce et Gao, 1996) : les coefficients de différentes résolutions sont donc directement comparables.

Afin de mieux comprendre les principes de la transformée en ondelettes, on représente sur la figure A.7 en haut à gauche, un exemple de signal discret 1D  $z$  de  $32 = 2^5$  valeurs puis sa transformée en ondelettes suivant 3 résolutions différentes et avec pour chaque série de coefficients, la même échelle sur l'ordonnée : de  $r = 4 = 2^2$  (8 translations) à  $r = 16 = 2^4 = 2^J$  (2 translations).

Chaque sous-figure sur la colonne de gauche représente les valeurs des coefficients  $S_{J,k}$  et  $d_{j,k}$ . Sur la deuxième colonne en partant de la gauche, les fonctions ondelettes dilatées et translatées  $\phi_{J,k}(t)$  (Eq. A.6) et  $\psi_{j,k}(t)$  (Eq. A.5) sont représentées chacune suivant une couleur différente. Sur la troisième colonne, la part du signal que l'on cherche à approcher avec les fonctions de la deuxième colonne correspond à la courbe épaisse (noire ou bleue). La courbe rouge correspond aux fonctions de la deuxième colonne multipliée par les coefficients d'ondelettes correspondants situés sur la première colonne. Sur la colonne de droite, on représente les résidus de l'approximation précédente de la courbe épaisse par la courbe rouge. Ces résidus deviennent le signal à approcher à la résolution inférieure suivante (ligne suivante sur la figure). Sur cet exemple,  $z(t)$  est quasiment totalement approché par les ondelettes aux deux résolutions les plus grossières  $r = 16, 8$ .

### A.2.2.2 Exemples virtuels de concavités et coefficients d'ondelettes de Haar associés

Un scalogramme (Bruce et Gao, 1996) est une image temps ( $k$ ) - fréquence ( $j$ ) ou résolution ( $r$ ) : chaque coefficient est représenté par la même surface sur cette image

---

<sup>2</sup> multiples de 2.

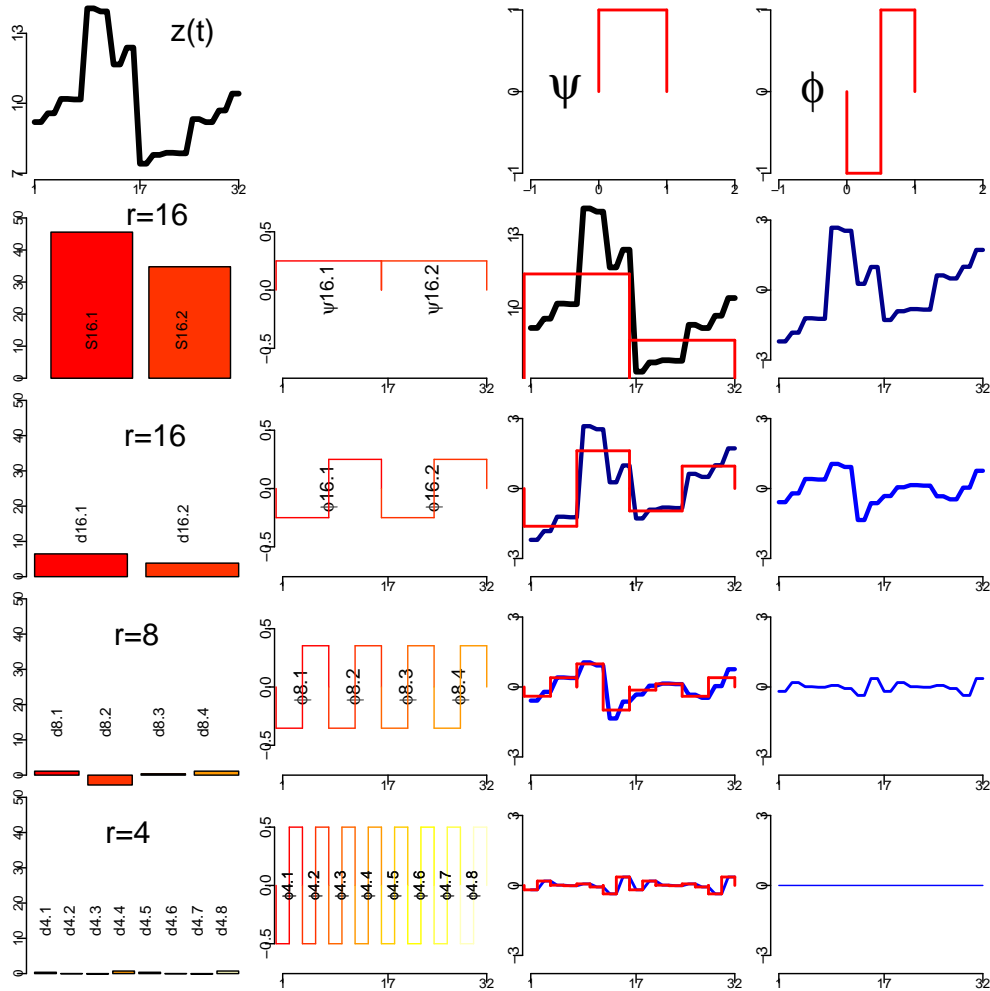


FIG. A.7 – Exemple d'un profil  $z(t)$  à 32 valeurs transformé par la paire d'ondelettes de Haar suivant 3 résolutions :  $r = 16, 8, 4$

(haute et étroite pour les résolutions les plus fines, basse et large pour les résolutions les plus grossières), suivant un niveau de gris relatif à sa valeur. Les coefficients de la même résolution sont successivement placés sur une ligne du scalogramme suivant l'ordre des coefficients de translation  $k$ .

La figure A.8 représente 3 séries de profils altimétriques virtuels (32 valeurs). Ces profils ont soit une concavité, soit une convexité en leurs centres (17ème valeur). Les scalogrammes de ces profils montrent la manière dont les coefficients d'ondelettes de détail sont reliés aux concavités et convexités. Ils montrent également que ces coefficients peuvent être des variables synthétiques discriminantes pour la détection et caractérisation de discontinuités topographiques comme les fossés :

- une concavité centrale induit une série successive d'un coefficient négatif (gris foncé) puis positif (gris clair) du fait de la forme de l'ondelette  $\phi$  de Haar ; A l'opposé, une convexité centrale induit une succession d'un coefficient positif

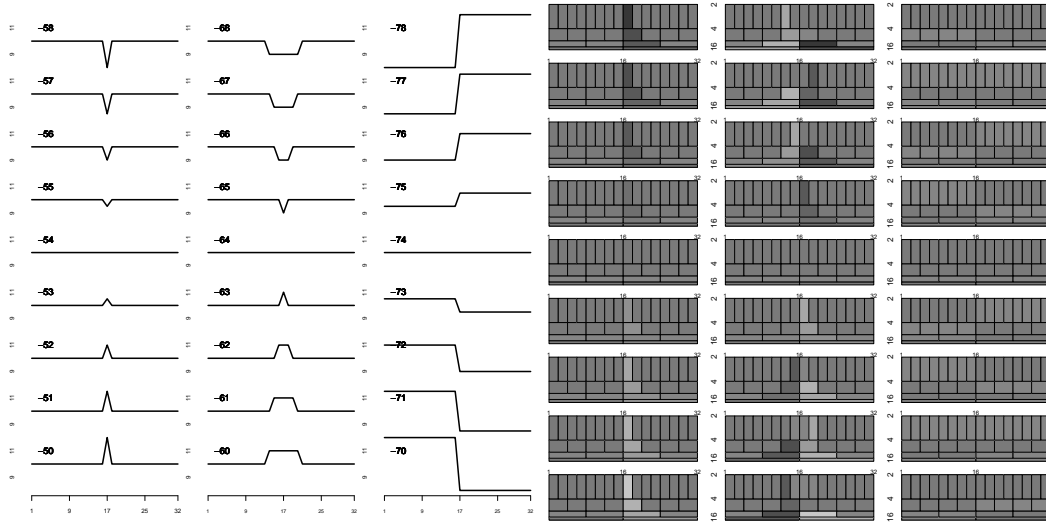


FIG. A.8 – Trois séries de profils topographiques virtuels et scalogrammes correspondants des coefficients de détail de l’ondelette de Haar  $d_{j,k}$  suivant une échelle de niveaux de gris allant de  $-2.75$  (noir) à  $2.75$  (blanc).

puis négatif. On peut noter également qu’une concavité centrale dans le signal induit une corrélation entre coefficients d’ondelettes de résolutions différentes mais identiquement localisés ;

- la valeur de la différence entre ces coefficients successifs est reliée à la profondeur de la concavité ou à la hauteur de la convexité (voir profils -50 à -58) ;
- la localisation de cette succession de valeurs positives-négatives dans les scalogrammes est liée à la localisation de la concavité dans le profils (ici, toujours en son milieu) ;
- la résolution  $r$  ayant la plus grande différence entre coefficients successifs traduisant la convexité ou concavité, est liée à la largeur de la concavité ou convexité, notamment lorsque  $r$  est proche de cette largeur (voir profils -60 à -68).
- l’allure générale du profil est capturée par  $\phi$  et  $\psi$  à la résolution la plus grossière  $r = 16$  (uniquement pour  $\phi$  ici).

### A.2.2.3 Comparaison de la transformée en ondelettes décimées (D.W.T) et continues (S.W.T)

Dans le cas d’une transformée en ondelettes continue (SWT : Stationary Wavelet Transform), à chaque échelle  $j$ , on obtient  $n$  coefficients pour un signal à  $n$  valeurs  $z = (z_1, \dots, z_n)$  ( $2^j$  coefficients pour DWT). SWT sur-échantillonne aux résolutions grossières par translation du signal. SWT est invariante en translation (d’où son nom de transformée en ondelettes stationnaire). Lorsque l’on transforme  $z$  par SWT suivant  $J$  échelles, on obtient  $(J + 1) * n$  coefficients au total, fortement corrélés entre eux (Bruce et Gao, 1996). De plus, d’autres hypothèses telle une périodicité du signal, doivent être introduites pour traiter (par translation) les "bords" du signal. SWT produit par conséquent des effets de bordure.

Afin de mieux comprendre les différences entre SWT et DWT, la figure A.9 représente 4 profils virtuels (les mêmes que ceux de la figure 4.5 p. 102) et leurs coefficients DWT

et SWT respectifs suivant 4 résolutions.

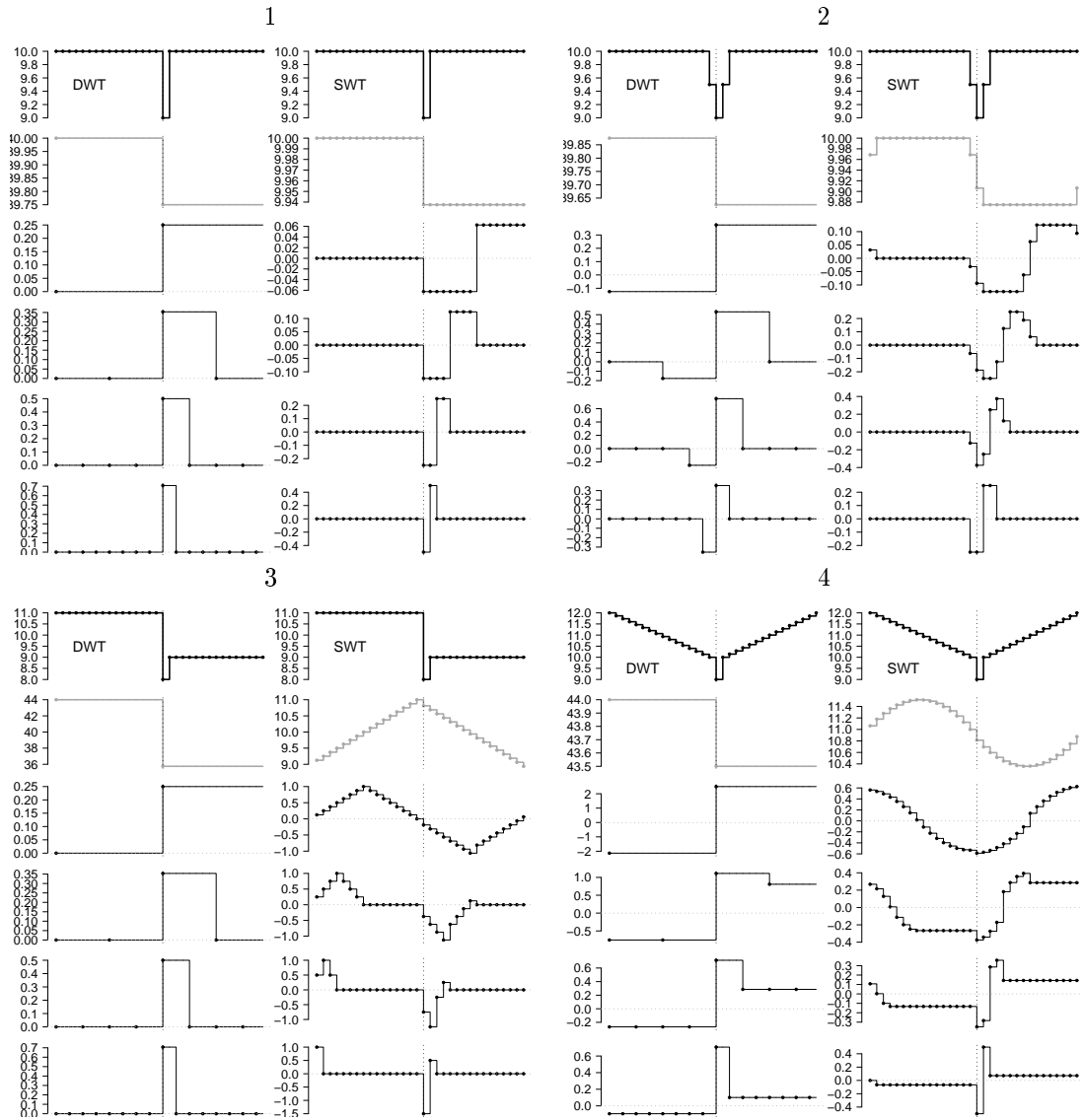


FIG. A.9 – Comparaison des coefficients DWT et SWT pour 4 profils topographiques virtuels ayant une concavité centrale marquée : les profils et coefficients D.W.T sont les mêmes que ceux de la figure 4.5 p. 102.

Sur cette figure, chacun des profils virtuels est représenté en haut et en gras. De haut en bas, on représente les coefficients de tendance  $S_{4,k}$  par les lignes et points gris et les coefficients de détail  $d_{j,k}$  de l'échelle la plus grossière ( $j = 4$ ) à l'échelle la plus fine ( $j = 1$ ) sont représentés par les lignes et points noirs. Les coefficients DWT sont sur la gauche et ceux issus de SWT sont sur la droite. Ces exemples de transformation montrent que :

- la tendance générale du signal est capturée par les coefficients de détail les plus grossiers  $d_{J,k}$  et par la différence entre les deux coefficients  $S_{J,k}$  pour DWT ;
- pour DWT comme SWT, une concavité centrale se traduit par une alternance de valeurs faibles-fortes ou négatives-positives des coefficients centraux ;
- SWT produit des effets de bordure comme on peut le voir sur les profils 3 and 4 ; la tendance du signal traduite dans les coefficients à échelle grossière est donc perturbée ;
- SWT produit également des coefficients de détail redondants, augmentant inutilement la dimensionnalité de l'espace de discrimination lorsque l'on souhaite utiliser ces coefficients comme variables prédictives de classifications de courbes (profils) ;
- la séquence de coefficients centraux  $d_{1,k}$  est la même pour DWT sur les profils 1 et 3 qui ont la même concavité centrale. Elles sont différentes dans le cas de SWT ;
- la valeur absolue de la différence dans la séquence de coefficients centraux suivant différentes résolutions est corrélée à la largeur de la concavité pour DWT (voir profils 1 and 2). C'est loin d'être le cas avec SWT.

Sur la base de ces exemples, on peut voir que lorsque l'on souhaite utiliser la transformée en ondelettes, non pas pour un problème de filtrage mais pour un problème de classification d'un signal  $1D$  discret régulier et relativement 'court', la transformée décimée DWT semble préférable.



### A.2.3 A.C.P sur les coefficients d'ondelettes : interprétation et sélection des facteurs

Une A.C.P avec normalisation est réalisée sur le tableau de 6461 individus et 13 variables, correspondant aux 13 coefficients d'ondelettes centraux sélectionnés :  $S_{4,1-2}$ ,  $d_{4,1}$ ,  $d_{4,2}$ ,  $d_{3,2}$ ,  $d_{3,3}$ ,  $d_{2,3}$ ,  $d_{2,4}$ ,  $d_{2,5}$ ,  $d_{2,6}$ ,  $d_{1,7}$ ,  $d_{1,8}$ ,  $d_{1,9}$ ,  $d_{1,10}$ .

#### A.2.3.1 Interprétation des composantes principales

A la lecture des valeurs propres de l'A.C.P, ces valeurs pour les troisième et quatrième facteurs PC3 et PC4 sont quasi-égales. Dans ce cas, seul le plan PC3-PC4 est à considérer, par rapport aux autres facteurs principaux. Comme la situation initiale de localisation des profils  $D$  par rapport aux  $ND$  dans ce plan PC3-PC4 n'est pas favorable pour CART (segmentations "obliques" : Fig. A.10), on réalise une rotation de repère dans ce plan PC3-PC4 pour construire de nouveaux facteurs  $PC3bis - PC4bis$ , orthogonaux entre eux (et aux autres). Ces nouveaux axes sont créés de manière à ce que l'axe  $PC3bis$  soit aligné sur la série de profils -50 à -58 de la figure A.8, signant le passage d'une convexité à une concavité centrale sur les profils.  $PC4bis$  est orthogonal à  $PC3bis$  (Fig. A.11). L'axe  $PC3bis$  devient une mesure directe de la profondeur des concavités centrales (ou des hauteurs de convexités centrales pour les valeurs négatives de  $PC3bis$ ).

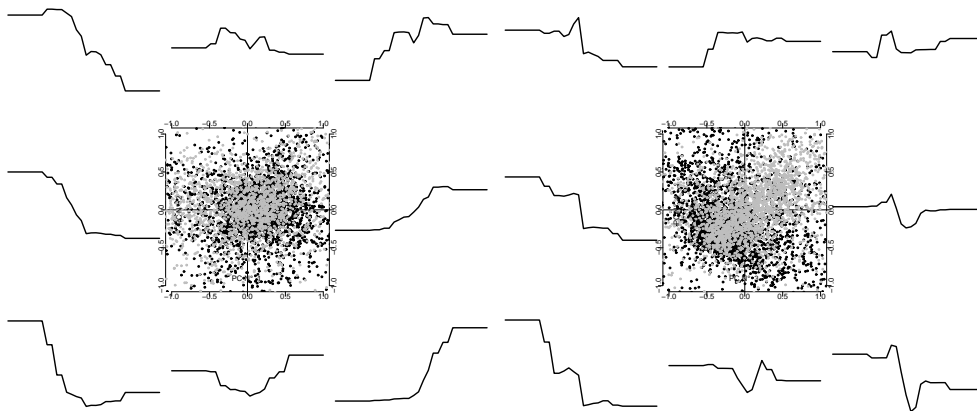


FIG. A.10 – profils  $D$  (points gris) et  $ND$  (points noirs) dans les plans des composantes principales - Gauche : plan  $PC1, PC2$ , Droite : plan  $PC3, PC4$ . Pour les profils aux valeurs de composante principale "extrêmes" ( $> 1$  ou  $< -1$ ), un profil "médian" est reconstruit (lignes épaisses noires), secteur par secteur, par inversion de l'équation 4.1 à partir des valeurs médianes des coefficients d'ondelettes.

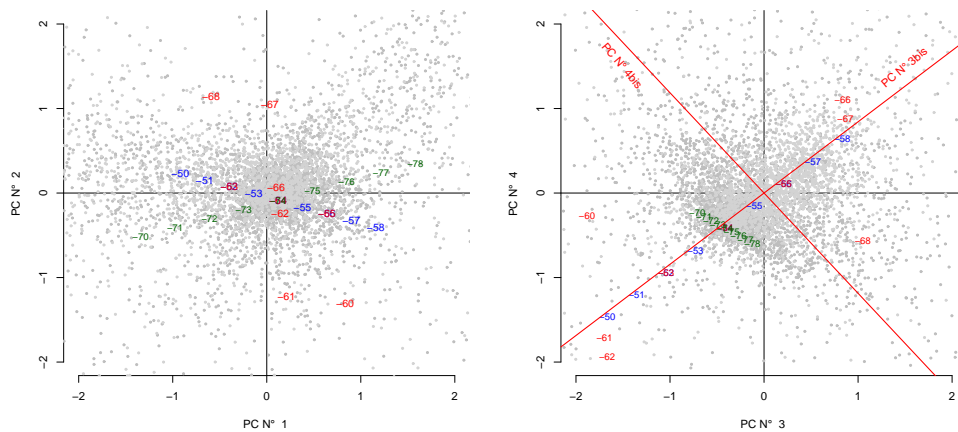


FIG. A.11 – Position des séries de profils topographiques virtuels (chiffres) et des 6461 profils (points gris) dans les plans de composantes avant rotation (gauche) et après (droite : nouveaux axes en rouge).

### A.2.3.2 Selection des composantes discriminantes

Suivant la procédure expliquée en page 108, la figure A.12 montre l'évolution des performances bootstrap de classification en ajoutant une à une les composantes principales et ce, en utilisant soit l'ensemble des composantes de départ ( $PC3 - PC4$ ) à gauche, soit l'ensemble des composantes de départ ( $PC3bis - PC4bis$ ) à droite. Si pour le taux global, il n'y a pas de différence, en revanche, la rotation dans le plan  $PC3 - PC4$  introduit des résultats moins dissymétriques entre les performances des profils  $D$  et  $ND$ . Dans les deux cas, on peut voir que le maximum est atteint dès 4 composantes (59 % d'inertie du tableau  $6461 * 13$  de départ).

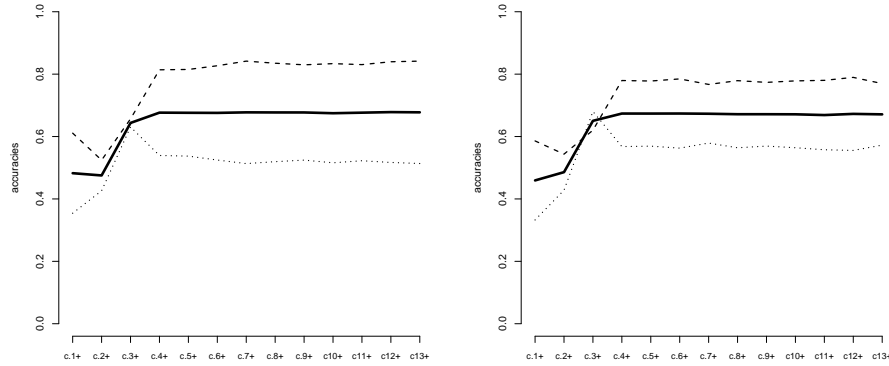


FIG. A.12 – Evolution des performances de classification globales (courbe en noir gras), pour  $D$  (...), pour  $ND$  (- -) en ajoutant une à une les composantes principales comme prédicteurs et avec 1-  $PC3 - PC4$  à gauche et 2-  $PC3bis - PC4bis$  à droite.

#### A.2.4 Effet du couple d'ondelettes sur les résultats de classification des profils altimétriques

Dans la table A.2, on présente des résultats de classification calculés suivant la méthode décrite page 108 sur les 6461 profils mais en utilisant différentes paires d'ondelettes courantes : ondelettes de Haar et de Daubechies, dont l'ondelette mère  $\psi$  est représentée.

$\psi$ shape	Wavelets	accuracy
	Haar	0.71
	Daubechies N.4	0.68
	Daubechies N.8	0.50
	Daubechies N.16	0.63

TAB. A.2 – Performances de classification en mode  $D, ND$  par utilisation de différentes paires d'ondelettes

On voit sur ce tableau que la paire de Haar est la plus adaptée à notre cas d'étude. Ceci est probablement dû au caractère compact et simple de cette paire, mieux adapté aux signaux de faibles longueurs.

## A.2.5 Algorithme CART

Dans cette annexe, nous présentons dans le détail les principes de l'algorithme d'apprentissage d'un arbre de classification CART, algorithme simple de segmentation proposé par Breiman *et al.* (1984). (voir <http://www.grappa.univ-lille3.fr/polys/apprentissage/sortie004.html>). Le principe général est de construire un arbre binaire de segmentation suivant un algorithme récursif dans une première phase dite d'expansion sur l'échantillon d'apprentissage puis de l'élaguer sur la base d'un échantillon test, exclusif de l'échantillon d'apprentissage.

### A.2.5.1 Expansion de l'arbre sur l'échantillon d'apprentissage

Soit  $S$ , l'ensemble des individus (profils)  $s_1, \dots, s_n$  d'apprentissage de classe  $Y = j$  avec  $j = \{0, 1\}$ . Soient  $X^1, \dots, X^P$  les variables descriptives ou prédictives.  $s_1$ , par exemple, est donc décrit par  $x_1^1, \dots, x_1^P$  et  $y_1$ .

On introduit une mesure d'impureté d'un sous-ensemble  $N_i$  de  $S$  (un noeud) qui mesure le degré de mélange entre les différentes classes  $j$ . Par défaut, l'indice de Gini est souvent utilisé. Il est défini par :

$$Gini(N_i) = 1 - \sum_{j=1}^J \frac{\text{card}(Y_{N_i} = j)^2}{\text{card}(Y_{N_i})^2}$$

Cet indice varie donc entre 0 et 0.5 dans le cas binaire. Il est minimal (pureté maximale) quand  $N_i$  contient une seule classe et est maximal (impureté maximale) quand les classes ont le même cardinal dans  $N_i$ .

L'algorithme de CART est le suivant :

1. Démarrer avec la racine  $N_i = S$  ;
2. Décider si  $N_i$  est un noeud terminal : c'est la cas si  $N_i$  est un noeud pur ( $Gini(N_i) = 0$ ) ou si il répond à un critère d'arrêt  $c$  qui limite la profondeur de l'arbre (par exemple,  $\text{card}(N_i) \leq c$ ). Dans ce cas, stopper la partition de  $N_i$ , sinon aller en 3 ;
3. Si  $N_i$  n'est pas un noeud terminal, chercher alors la partition binaire de  $N_i$  en  $N_{j1}$  et  $N_{j2}$  qui maximise une fonction de gain de pureté :
  - Réaliser pour chaque valeur seuil  $a_i^p$  des  $X_1, \dots, X_P$ <sup>3</sup>, la partition binaire de  $N_i$  en  $N_{j1}$  et  $N_{j2}$  :

$$N_{j1} = \{s \in N_i / x^p < a_i^p\} \text{ et } N_{j2} = \{s \in N_i / x^p \geq a_i^p\}$$

- calculer alors une fonction de gain de pureté pour cette partition :

$$g(N_i, a_i^p, X_p) = Gini(N_i) - \sum_{j=\{j1, j2\}} \frac{\text{card}(N_j)}{\text{card}(N_i)} * Gini(N_j)$$

---

<sup>3</sup>si  $X_p$  est qualitative à  $k$  modalités cela représente  $2^{k-1}$  possibilités; si  $X_p$  est quantitative, alors on réalise un nombre fini de tests basés sur une discretisation de  $X_i$

– Sélectionner le seuil  $a$  et la variable  $X_k$  qui maximise cette fonction de gain :

$$(a, X_k) = \max_{a_l^p, X_l \in X_1, \dots, X_p} g(N_i, a_l^p, X_p)$$

4. Pour chaque noeud généré  $N_{j1}, N_{j2}$  suivant la partition  $(a, X_i)$  sélectionnée, repartir au point 2.

Ainsi, pour chaque feuille terminale formée, on affecte une classe qui est la classe majoritaire.

#### A.2.5.2 Élagage de l'arbre sur un échantillon test

Soit  $G_0$ , l'arbre obtenu à la fin de la phase d'expansion. On construit alors une suite d'arbres

$G_0, G_1, \dots, G_i, G_{i+1}, \dots, G_q$  telle que  $G_{i+1}$  est un élagué de  $G_i$  de un noeud. Pour ce faire, il nous faut définir le procédé de construction de  $G_i$  à  $G_{i+1}$ .

Ce procédé passe par le calcul d'une quantité, pour chaque sous-arbre  $G_i^r$  de  $G_i$  élagué d'un noeud  $r$ , liée à la variation d'erreur apparente  $\Delta_r$  sur l'échantillon d'apprentissage. On note cette quantité :

$$g(r) = \frac{\Delta_r}{o - 2}$$

avec  $o$ , taille de  $G_i$  et  $\Delta_r$ , variation d'erreur apparente entre  $G_i$  et  $G_i^r$  pour les échantillons d'apprentissage formant un ensemble  $A_r$  situé au noeud  $r$ . On peut noter :

$$\Delta_r = \frac{MC_{G_i}(r) - MC_{G_i^r}(r)}{\text{card}\{A_r\}}$$

avec  $MC_{G_i}(r)$ , le nombre d'échantillons d'apprentissage mal classés au noeud  $r$  suivant  $G_i$  et  $MC_{G_i^r}(r)$ , le nombre d'échantillons d'apprentissage mal classés au noeud  $r$  suivant  $G_i^r$ .

On sélectionne ensuite  $G_{i+1} = G_i^r$  qui minimise la fonction  $g(r)$ .

Puis, on calcule pour la suite  $G_0, G_1, \dots, G_i, G_{i+1}, \dots, G_q$ , les erreurs apparentes (mal classés) sur l'échantillon test. On sélectionne ensuite dans cette suite l'arbre qui minimise cette erreur apparente.

## A.2.6 Comparaison de performances de classification de profils altimétriques LiDAR suivant les algorithmes CART et RANDOM FOREST

### A.2.6.1 Introduction

L'objectif de cette annexe est de comparer les performances de classification suivant un mode binaire **D**, **ND** (fossé, non-fossé) de profils altimétriques LiDAR transformés en ondelettes suivant deux méthodes de classification. Ces méthodes de classification sont liées entre elles et font toutes deux intervenir des rééchantillonnages aléatoires afin de lisser les résultats d'un modèle lié à une échantillonnage. Il s'agit :

- d'un arbre de classification (**CART**) sélectionné suivant la méthode présentée en page 106;
- et sa déclinaison par randomisation multiple connue sous le terme de forêt aléatoire (forêt, i.e. ensemble de plusieurs arbres) ou "Random Forest" (**RF**).

### A.2.6.2 Données

Pour mener cette comparaison, on utilise successivement les sites **B** et **C**, comprenant une diversité d'unités morphologiques. Le site **B**, d'une surface de  $2 \text{ km}^2$ , contient 6461 individus (profils) : c'est le site qui a fait l'objet des développements méthodologiques présentés en section 4.2.1 ; ce site sera principalement utilisé pour la calibration des modèles ainsi que pour réaliser des premières validations croisées. Le deuxième site, le site **C**, d'une surface de  $1 \text{ km}^2$  contient 2999 profils à classer : il est utilisé pour un deuxième niveau de validation croisée permettant de juger plus avant de la robustesse des modèles de classification CART et RF. Les variables discriminantes sélectionnées pour ces ensembles sont les quatre premières composantes (avec rotation) issues du tableau des coefficients d'ondelettes 'centraux', résultants de la transformée des profils (voir section 4.2.1).

### A.2.6.3 Méthodes

**A.2.6.3.1 Arbre de classification CART sélectionné par rééchantillonnage aléatoire** On reprend exactement la méthode présentée en page 106. On effectue ici 500 tirages aléatoires.

**A.2.6.3.2 Random Forest** CART est une méthode de segmentation de l'espace des classes dans l'espace des variables un peu brutale et très sensible à l'échantillonnage car non paramétrique. Pour pallier à cet inconvénient, cette méthode a évolué vers des dérivés "randomisés" et "agrégés" selon deux directions :

- une première a donné naissance au 'bagging' pour 'bootstrap aggregating' (Breiman, 1996a). L'idée du bagging est de construire non pas un mais un ensemble d'arbres, sans élagage, par rééchantillonnage bootstrap de l'ensemble de calibration (comme précédemment pour CART) puis de retenir in fine pour un individu à classer le vote majoritaire issu de la combinaison de tous les arbres construits.
- d'autres randomisations sur CART ont été proposées en parallèle. Notamment une randomisation dans la procédure de construction de l'arbre ou à chaque noeud, plutôt que de retenir la variable qui maximise le critère de gain de pureté, on en sélectionne une aléatoirement parmi les  $K$  premières ou même une

seule aléatoirement parmi toutes les possibles. On peut également ne considérer comme variables candidates pour la partition binaire d'un noeud qu'un sous-ensemble aléatoire des variables, en général,  $\sqrt{P}$ , avec  $P$  nombre total de variables. De même, en répétant cette procédure puis en faisant voter chaque arbre construit, on peut retenir le vote majoritaire. RF est une généralisation des deux randomisations évoquées. Cette méthode s'est montrée plus performante que d'autres méthodes issues du data mining et d'arbres de segmentations comme le boosting (méthodes de pondération des échantillons lors de la construction de l'arbre) (Breiman, 1996b). Elle est surtout utilisée lorsque l'on a beaucoup de variables prédictives, car elle propose, dans ses résultats des outils de sélection de variables discriminantes.

Pour l'utilisation de la méthode RF, on effectue l'échantillonnage aléatoire suivant :

A partir du même ensemble de 6461 profils, 1– on en sélectionne 1686 D et 1686 ND sans remise (il n'y a plus d'échantillon test d'élagage) pour la calibration de la forêt et on valide les résultats sur 845 D et 845 ND exclusifs. A chaque noeud, les  $\sqrt{4} = 2$  variables candidates à la partition binaire sont tirées aléatoirement parmi les 4. De même que pour CART, on répète cette procédure 500 fois. On obtient donc 500 matrices de confusion. En revanche, contrairement à CART, la segmentation finale ne peut être visualisée puisqu'elle est la combinaison de nombreux arbres (pas d'interprétation du modèle de classification possible).

#### A.2.6.4 Résultats

**A.2.6.4.1 Sur le site B** La figure A.13 montre la distribution des éléments de la diagonale des matrices de confusion, i.e. successivement les performances de classification globale, puis pour  $D$  en enfin pour  $ND$ . Les distributions sont en bleu pour RF et en rouge pour CART. Ces taux sont issus de validations croisées sur 1590 échantillons du site B. Les résultats moyens sont détaillés ci-après.

Pour CART, on obtient :

- un taux global de 67 %,
- un taux de 57 % pour les D,
- un taux de 78 % pour les ND.

Pour RF, on obtient :

- un taux global de 77 %,
- un taux de 80 % pour les D,
- un taux de 73 % pour les ND.

RF est plus performant globalement et plus stable : on a une moindre variance des taux de classement (logique car l'échantillon d'apprentissage est deux fois plus important, on explore donc moins de variabilité d'échantillonnage) et un équilibre des performances par classe qui se traduit par une augmentation très significative du taux de classement moyen des  $D$ .

Appliqué à l'ensemble du site B, on obtient les résultats suivants : Pour CART :

- un taux global de 71 %,
- un taux de 54 % pour les D,
- un taux de 83 % pour les ND.

Pour RF :

- un taux global de 83 %,
- un taux de 84 % pour les D,

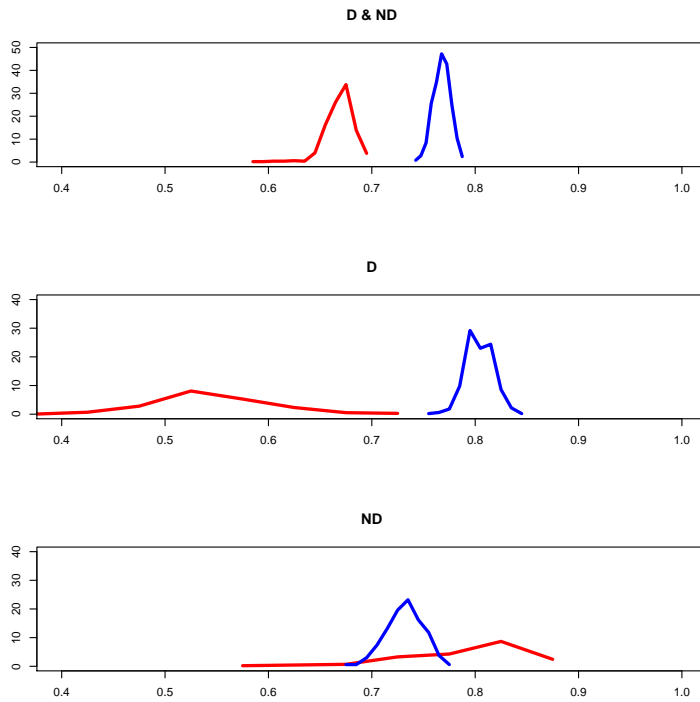


FIG. A.13 – Distribution des taux de bien classés sur  $(D + ND)$ ,  $D$ ,  $ND$  de haut en bas; Bleu : **CART** ; Rouge : **RF**.

- un taux de 82 % pour les ND.

La comparaison des résultats cartographiés en figure A.14 montre dans les deux cas un fort taux de commission de fossé sur les zones de coteaux et végétalisées.

**A.2.6.4.2 Sur le site C** Si on applique les modèles CART moyen et RF obtenus précédemment sur tous les échantillons de le site **C**, ce nouveau test de validation montre des résultats globaux similaires entre les deux méthodes, mais différemment répartis. En effet, on obtient :

- un taux identique de bien classés de 72 %,
- qui se décline en 87% contre 74% pour les ND entre CART et RF,
- et 56 % contre 71 % pour les D entre CART et RF.

Les cartes comparées des résultats de classification sur cette zone (Fig. A.15) montrent des distributions spatiales relativement proches entre les deux méthodes si ce n'est pour la zone de coteaux en friche et forêt qui est très mal classée par RF, du fait d'un nombre de sur-détections important.





FIG. A.14 – Résultats cartographiés des modèles moyens CART et RF sur le site B : les points bleus sont les  $D$  bien classés, les rouges les  $D$  omis, les orange les  $ND$  classés en  $D$  et les noirs les  $ND$  bien classés.

#### A.2.6.5 Conclusion

La méthode de classification RF ne semble pas globalement plus performante que le modèle CART moyen lors d'une extrapolation du modèle de classification sur une zone de validation, extérieure à la zone d'apprentissage. Les bons résultats de RF sur le site **B** (calibration) peuvent s'expliquer par le fait que :

1. on a doublé l'échantillonnage d'apprentissage, d'où une moindre variance des performances due au rééchantillonnage. On améliore uniquement l'erreur de calibration.
2. il existe une corrélation spatiale entre individus (des profils LiDAR voisins se ressemblent), et donc les tests par validation croisée sur les 1590 individus du site **B** sont d'autant plus sensibles à l'importance du "sur-ajustement" sur cette zone, ce qui semble le cas de la méthode RF.
3. On obtient également un modèle RF plus équilibré en terme de matrice de confusions sur les classes  $D$  et  $ND$ . Ce résultat va dans le sens de critères de

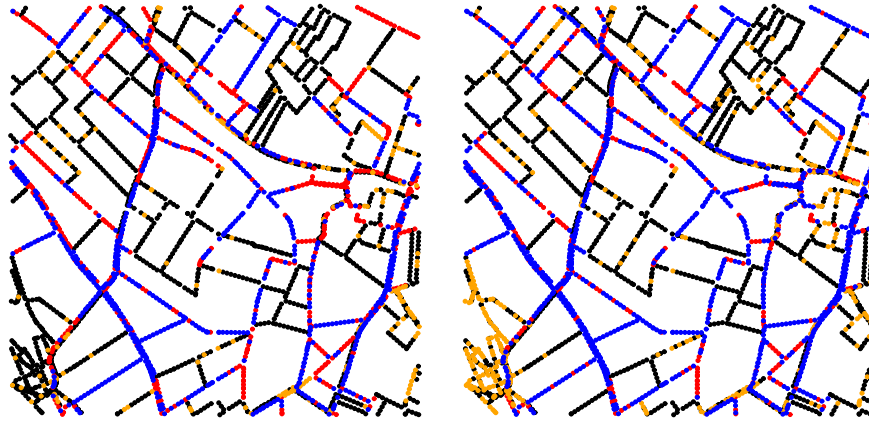


FIG. A.15 – Résultats cartographiés des modèles moyens **CART** et **RF** sur le site **C** : les points bleus sont les  $D$  bien classés, les rouges les  $D$  omis, les orange les  $ND$  classés en  $D$  et les noirs les  $ND$  bien classés.

classification uniquement statistiques liés à la composition de l'échantillonnage de calibration pour RF, qui lisse fortement les dissymétries de la méthode CART. Cette meilleure détection des fossés est compensée par une très forte sur-détection et donc une localisation de fossé "homogénéisée" dans l'espace, ce qui n'est peut-être pas souhaitable.

## A.2.7 Détection de fossés par classification de profils radiométriques issus d'images optiques à très haute résolution spatiale

Cette annexe est extraite d'une présentation orale réalisée au colloque *EGU* de Vienne en 2006, intitulée : "Comparison of artificial drainage network detection rates in cultivated landscapes with LiDAR and high resolution CIR images" (Bailly *et al.*, 2006a).

### A.2.7.1 Methods

Methodology to detect ditches presence on cadastral lattice was exactly the same as the one used with LiDAR altimetric data. We first built radiometric profiles on the same 6461 sites from both **I1** and **I2** images on the **B** study area. We then transformed profiles with D.W.T and classified profiles starting from the 13 central wavelet coefficients since a ditch on an image creates a radiometric concavity due to shadows. The methodology for dimension reduction, classification process and classification model selection was the same as with LiDAR data.

### A.2.7.2 Images data

**A.2.7.2.1 CIR panchromatic image I1** Simultaneously to LiDAR data, CIR images having 0.5 m spatial resolution were acquired on **B** study area and geometrically corrected using LiDAR DTM. We transformed the multi-spectral CIR image to a panchromatic one, in grey levels, while linear shadows corresponding to ditches are informative on grey levels images (Fig. A.16).

**A.2.7.2.2 CIR panchromatic image I2** CIR images were acquired on **B** study area in december 2004 with a non-metric camera and EIR films from an ultra light plane. Climatic conditions were very dry before and during this period. These images were scanned and joined in a mosaic with a geometric reference coming from the CIR **I1** image. Plot-boundaries x,y positioning RMS<sup>4</sup> was computed comparing to geographical database : this was estimated to 1 m. As with summer image, we transformed multi-spectral CIR image to panchromatic one in grey levels (Fig. A.16) with a 0.5 meter spatial resolution.

For both panchromatic **I1** and **I2** images, radiometries were not equalized all over test areas. This was not done while only local contrasts on greys levels are used further for ditches detection or classification. We did not chose to transform CIR image in NDVI<sup>5</sup> while local contrasts at ditches locations on NDVI were weaker.

### A.2.7.3 Visual interpretation

Radiometric profiles are called further the *I1* profiles for the summer ones and *I2* for winter ones. Visual classifications of each profile on sites *D* on panchromatic images were performed in two classes (with ditches detection or not) : ditches on panchromatic images correspond to linear shadowed elements coming from local concavities in the terrain when any vegetation is masking it. Detection rates per morphological type are given in table A.3. This results show that winter image with

---

<sup>4</sup>Root mean square error

<sup>5</sup>Normalized Difference Vegetation Index



FIG. A.16 – Panchromatic images subsets on **B** study area for **I1** (left) and **I2** (right).

both lower sun elevation giving contrasted shadows in surface concavities, and less vegetation due to 1- an higher ditches maintain by farmers during winter and 2- cold and dry conditions during winter is more favorable for visual ditches detection. This is particularly true in flat and intensively cultivated areas in "depression" morphological units. These result show too that visual detection rates on LiDAR data and on winter panchromatic images are quasi equal.

Morphology	I2	I1
Mean	<b>0.42</b>	<b>0.27</b>
Depression	0.62	0.36
Glacis and terraces	0.32	0.29
Terrassettes	0.29	0.01
Tables	0.11	0.04

TAB. A.3 – Results from visual interpretation of linear concavities on *D* sites

#### A.2.7.4 Automated classification

Table A.4 shows profiles classification accuracies when applying the selected CART mean model to the 6461 *I2* or *I1* profiles. The better detection is obtained with LiDAR data (voir section 4.2.2), then with **I2** image then **I1**, in the same order as the according visual detection rates (Tab A.3).

Profiles	all D	all ND	all mean
<b>I2</b>	<b>0.52</b>	0.81	<b>0.69</b>
<b>I1</b>	<b>0.37</b>	0.82	<b>0.65</b>

TAB. A.4 – Profiles classification accuracies on **B** study area when using panchromatic images.

#### **A.2.7.5 Conclusion**

In this study, we assessed ditches detection rates we can expect when using high spatial resolution images. We compared these results to the ones obtained with LiDAR data on same sites and with the same methodology. When detection is automatically performed with radiometric profile classification using wavelet transform, we showed that ditches detection rates are conform to the ones obtained when classifying visually the profiles. Moreover, these results show that : 1- for same date acquisition, detection from LiDAR is higher to the one from image, 2- images with low sun elevation creating clear linear shadows along ditches gives interesting ditches detection capacities.

### A.2.8 Classification binaire des arêtes du lattice cadastral à partir des profils LiDAR classés suivant des modèles "dissymétriques"

La même procédure de sélection de règles de "lissage" que celle employée en sous-section page 116 est répétée en prenant les résultats de deux autres modèles de classification de profils LiDAR (à partir des transformées en ondelettes des profils) : un modèle dénommé *extND* qui présente peu de commissions de profils *D*, un autre dénommé *extD* avec, au contraire, peu d'omissions de profils *D* mais beaucoup de sur-détections de profils *D*.

Comme précédemment, on construit puis élague un arbre qui traduit la règle de lissage : On dénomme ARBRE-EXTD l'arbre construit sur les résultats *extD*, et ARBRE-EXTND, celui construit sur les résultats *extND*. Pour ARBRE-EXTND, la règle d'affectation des arêtes *D* est :  $Suite1max \geq 1$ ; Pour ARBRE-EXTD, la règle d'affectation des arêtes *D* est plus complexe mais s'appuie principalement sur la segmentation :  $Suite1max \geq 1$ . Ces arbres sont présentés ci dessous (ainsi que leurs résultats cartographiés). Dans le tableau A.2.8, on récapitule les taux de linéaires d'arêtes du lattice bien classées pour les deux zones (**B** et **C**), avec ces nouvelles règles de lissage, puis avec celle proposée précédemment dans le texte du manuscrit p. 116 (Eq. 4.2) dénommée ARBRE-MEAN, ou par simple règle majoritaire.

Sur ces résultats, on peut voir que :

1. Globalement, les profils *ext - D* donnent toujours des résultats plus faibles. Le modèle de lissage paraît sur-ajusté et les résultats difficilement interprétables.
2. Pour les profils *ext - ND*, la règle de lissage est assez proche de l'équation 4.2. D'ailleurs, les taux de linéaire d'arêtes classées suivant l'équation 4.2 appliquée aux profils *extND* sont assez proches en moyenne pour les deux zones de ceux obtenus par application de la règle ARBRE-EXTND appliquée aux mêmes profils *extND*. Ces taux sont également proches d'une simple agrégation suivant la classe majoritaire.
3. La règle de l'équation 4.2 semble "robuste", car, quelque soit l'origine des profils classés, elle améliore les taux de linéaire classé, et ce d'une manière assez équilibrée entre les zones **B** et **C**.

#### A.2.8.1 Lissage des résultats de classification de profils EXT-ND

La figure A.17 représente en haut, le modèle ARBRE-EXTND de lissage construit puis sélectionné par élagage sur les zones **B** et **C**, lorsque l'on part des résultats de classification de profils *extND*, i.e. avec peu de commissions de fossés.

#### A.2.8.2 Lissage des résultats de classification de profils EXT-D

La figure A.18 représente en haut, le modèle ARBRE-EXTD de lissage construit puis sélectionné par élagage sur les zones **B** et **C**, lorsque l'on part des résultats de classification de profils *extD*, i.e. avec peu d'omissions de fossés.

	$D + ND$	$D$	$ND$
	<b>Profils <i>ext</i> – <math>D</math></b>		
zone <b>B</b> avant lissage	<b>65</b>	67	64
zone <b>C</b> avant lissage	<b>69</b>	73	66
<b>lissage ARBRE-EXTD</b>			
zone <b>B</b> après lissage	<b>71</b>	47	86
zone <b>C</b> après lissage	<b>74</b>	51	90
<b>lissage ARBRE-MEAN</b>			
zone <b>B</b> après lissage	<b>66</b>	71	63
zone <b>C</b> après lissage	<b>71</b>	68	73
<b>lissage majoritaire</b>			
zone <b>B</b> après lissage	<b>69</b>	76	54
zone <b>C</b> après lissage	<b>67</b>	77	59
	<b>Profils <i>ext</i> – <math>ND</math></b>		
zone <b>B</b> avant lissage	<b>71</b>	39	92
zone <b>C</b> avant lissage	<b>71</b>	47	92
<b>lissage ARBRE-EXTND</b>			
zone <b>B</b> après lissage	<b>77</b>	54	90
zone <b>C</b> après lissage	<b>72</b>	48	88
<b>lissage ARBRE-MEAN</b>			
zone <b>B</b> après lissage	<b>74</b>	41	95
zone <b>C</b> après lissage	<b>73</b>	40	96
<b>lissage majoritaire</b>			
zone <b>B</b> après lissage	<b>74</b>	37	97
zone <b>C</b> après lissage	<b>73</b>	46	92

TAB. A.5 – Pourcentages de linéaire des arêtes du lattices bien classées avant et après lissage sur les zones **B** (calibration) et **C** (validation) par application de différentes règles de lissage sur différents résultats de classification des profils.

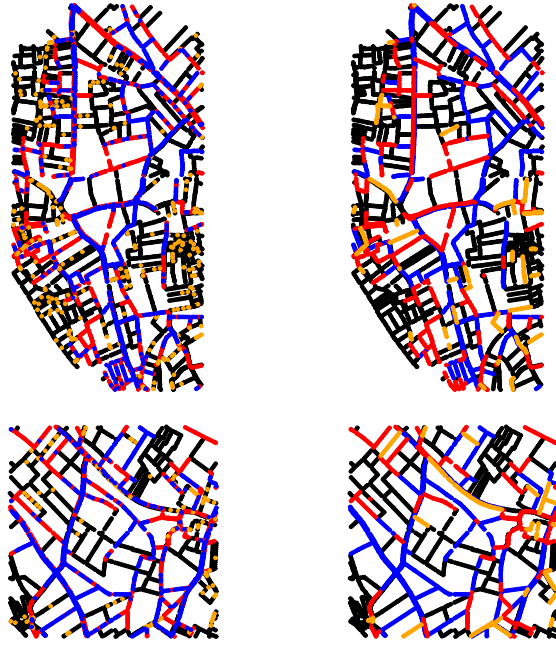
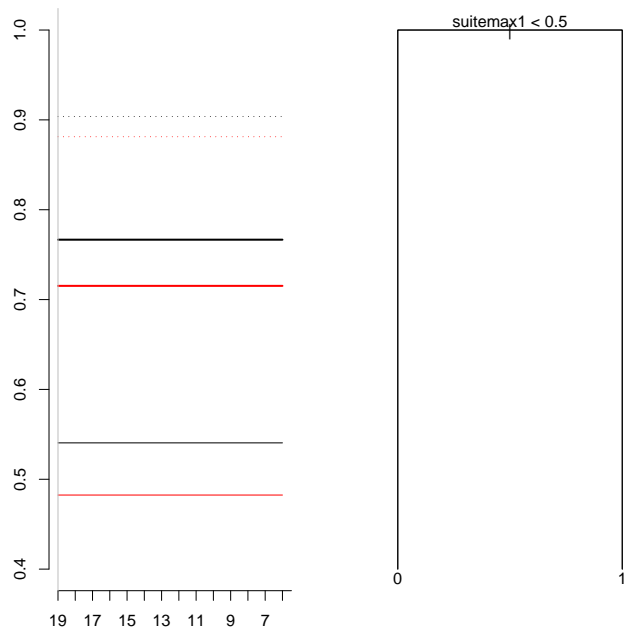


FIG. A.17 – Haut : Evaluation du linéaire d'arêtes bien classées suivant la complexité de l'arbre (Gauche) ; Arbre à 2 feuilles sélectionné (Droite). Bas : Comparaison des erreurs de classification avant et après lissage sur la zone **B**-haut et **C**-bas (même signification des codes couleurs).



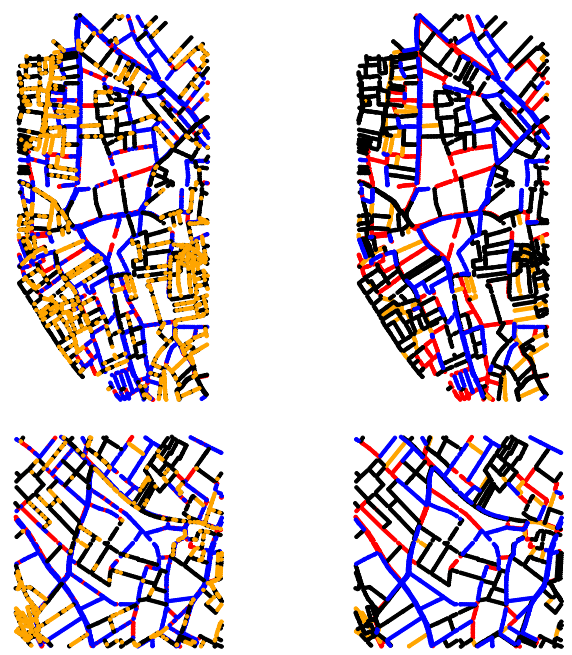
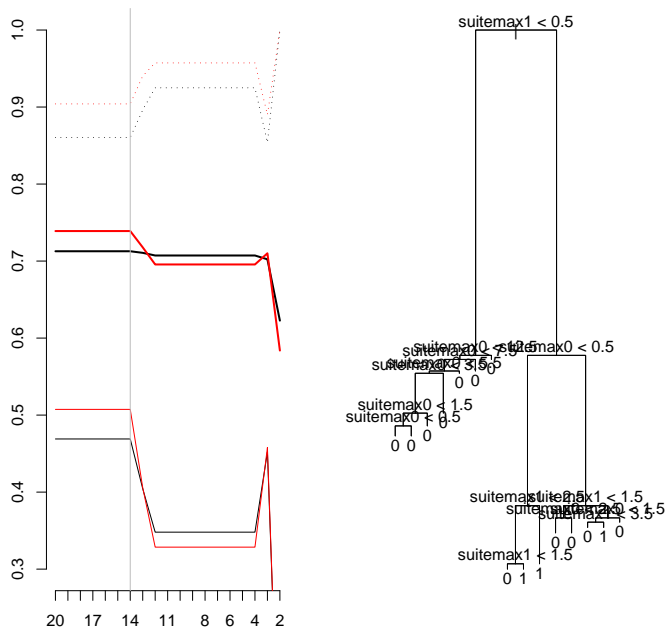


FIG. A.18 – Haut : Evaluation du linéaire d'arêtes bien classées suivant la complexité de l'arbre (Gauche) ; Arbre à 14 feuilles sélectionné (Droite). Bas : Comparaison des erreurs de classification avant et après lissage sur la zone **B**-haut et **C**-bas (même signification des codes couleurs).

### A.2.9 Loi de Tokunaga

L'idée des lois de Tokunaga est d'introduire une loi sur la statistique des arêtes dites "latérales". Les arêtes latérales désignent des arêtes se jetant dans des arêtes d'ordre de Horton-Stahler  $w$  supérieur, lesquels étant les arêtes absorbantes.

Le principe est le suivant : On compte pour un réseau donné le nombre moyen d'arêtes latérales d'ordre  $\nu$  se jetant dans une arête d'ordre  $\mu$  avec  $\mu > \nu$ . On note cette statistique, le ratio de Tokunaga  $T_{\mu,\nu}$ . On obtient donc une matrice triangulaire inférieure de ces ratios pour un réseau donné ou un graphe comme ceux présentés en figure A.19.

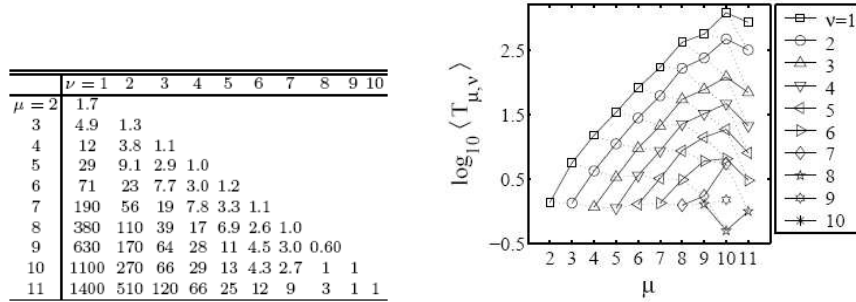


FIG. A.19 – Statistiques de Tokunaga sur le bassin du Mississippi en tableau (gauche) ou graphes (droite) (Dodds et Rothman, 1999)

De même par observation d'invariance d'échelle pour de grands réseaux hydrographiques, Tokunaga en a déduit une loi à deux paramètres  $(T_1, R_T)$  qui s'écrit, pour tout  $k = \mu - \nu$  :

$$\bar{T}_k = T_1 * R_T^{k-1} \quad (A.8)$$

Le paramètre  $T_1 > 0$  est le nombre moyen de segments latéraux se jetant dans un segment absorbant d'ordre supérieur. Le paramètre  $R_T$  mesure l'accroissement de la densité des segments latéraux lorsque leur ordre diminue. Ce paramètre est lié aux lois d'invariance d'échelle de Horton présentées auparavant. On peut trouver une démonstration sur cette liaison dans Dodds et Rothman (1999).

### A.2.10 Exemples de lois géométriques de longueurs des branches

Les lois de longueurs des arêtes de l'arborescence (ici prises comme segments de réseau entre deux confluences ou entre une source ou la racine et une confluence) sont fonction de l'ordre  $w$  de cette arête (Mendes-France, 1981). Sur la figure A.20, on peut visualiser trois exemples de lois géométriques de modèle général :

$$L(u) = \alpha_1 + \alpha_2 * w + \alpha_3 * w^2$$

La première (Gauche) est une loi quadratique, la deuxième, une loi linéaire ( $\alpha_3 = 0$ ) et la troisième (Droite), une loi constante ( $\alpha_3, \alpha_2 = 0$ ).

Certaines de ces lois (constante et linéaire) correspondent à des lois d'invariance de Horton des longueurs moyennes des segments d'ordre  $w$  (Dodds et Rothman, 1999).

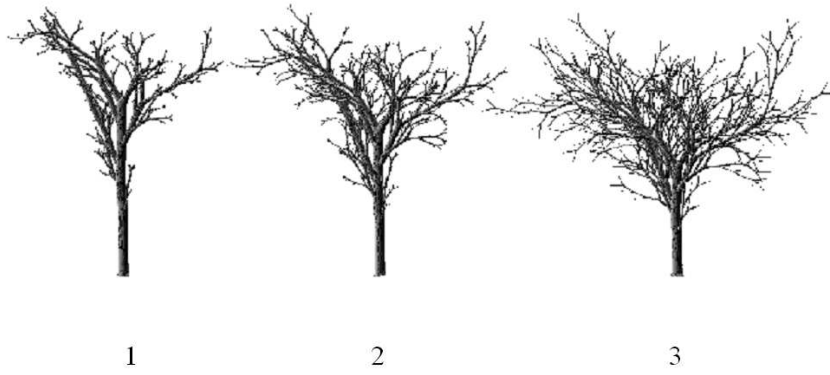


FIG. A.20 – Exemples d’arbres binaires (en 3D) suivant une loi de longueur de segments quadratique (1), linéaire (2), ou constante (3), extrait de (Janey, 1992).

# Spatial conditional simulation of directed trees : application to artificial drainage network of vineyard landscapes

---

## ABSTRACT

Artificial drainage networks impact strongly cultivated landscapes hydrology. The improvement of hydrological tools and methods that aim to assess the hydrological impacts of those artificial networks is dependent to our capabilities to map these networks, i.e. its directed tree architectures and its local characteristics, highly variable in space. This thesis develops stochastic methods that aim to map drainage networks. We first developed a geostatistical framework to simulate variabilities along directed trees. A method that simulate network architecture by linking network segments detected with LiDAR is secondly proposed. These developments simulate hydrographic networks in cultivated landscapes that are both conditioned by spatial sampling on networks and by the structure of the space that supports these simulations. These hydrographic networks conditional simulation methods are useful to complete scarce data and to asses hydrological model sensitivity to networks spatial structures.

---

## KEY WORDS

Geostatistic, LiDAR, Interpolation, Ditches, Uncertainties

---

## RÉSUMÉ

Sur les bassins versants cultivés, les aménagements qui guident les écoulements de surface concentrés et forment ainsi le réseau hydrographique influent fortement sur le régime et la qualité des eaux. Pour progresser sur les méthodes et outils qui permettent de quantifier les impacts de ces aménagements à l'échelle du bassin versant, la spatialisaiton de ces réseaux structurés en arborescences, de leurs propriétés locales, de leur architecture, fortement hétérogènes dans le paysage, est primordiale. Le travail de thèse réalisé porte sur le développement de méthodes de spatialisation probabilistes de ces réseaux. On développe des méthodes géostatistiques qui permettent de simuler des variabilités le long d'arborescences. Une méthodologie de simulation de l'architecture d'un réseau par connexion de segments du réseau détectés par télédétection LiDAR est ensuite proposée. Ces développements méthodologiques produisent des simulations de réseaux dans le paysage conditionnées par un échantillonnage spatial du réseau et par la structure de l'espace qui supporte ces simulations. Ces méthodes de simulations conditionnelles de réseaux hydrographiques artificiels sont utiles pour remplacer les données manquantes, tester des protocoles d'échantillonnage et réaliser des études de sensibilité de modèles hydrologiques sur un critère de structure spatiale de réseaux.

---

## MOTS CLÉS

Géostatistique, LiDAR, Interpolation, Fossés, Incertitudes

---

## DISCIPLINE

Sciences de la terre et de l'eau, CNU 35-36

---

## INTITULÉS ET ADRESSES DES LABORATOIRES D'ACCUEIL

### UMR LISAH

Laboratoire d'étude des Interactions  
Sol-Agrosystème-Hydrosystème



2, place Pierre Viala  
34060 Montpellier cedex1  
FRANCE

<http://sol.ensam.inra.fr/lisah/>

### UMR TETIS

Territoires, Environnement, Télédétection  
et Information Spatiale



500, rue J.F. Breton  
34093 Montpellier cedex5  
FRANCE

<http://www.teledetection.fr/>