



**HAL**  
open science

# Paysages virtuels et analyse de scénarios pour évaluer les impacts environnementaux des systèmes de production agricole

Luc Sorel

► **To cite this version:**

Luc Sorel. Paysages virtuels et analyse de scénarios pour évaluer les impacts environnementaux des systèmes de production agricole. Sciences de la Terre. Agrocampus - Ecole nationale supérieure d'agronomie de rennes, 2008. Français. NNT: . tel-00440205

**HAL Id: tel-00440205**

**<https://theses.hal.science/tel-00440205>**

Submitted on 9 Dec 2009

**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.

# Thèse de Doctorat

présentée devant  
AGROCAMPUS OUEST

Pour obtenir le titre de Docteur d'Agrocampus Ouest  
École doctorale Vie-Agro-Santé  
Mention : sciences de l'environnement

par

**Luc SOREL**

## **Paysages virtuels et analyse de scénarios pour évaluer les impacts environnementaux des systèmes de production agricole**

### **Virtual landscapes and scenario analysis to assess the environmental impacts of agrosystems**

Préparée à l'UMR 1069 SAS (Sol, Agro- et hydrosystèmes, Spatialisation)  
INRA / Agrocampus Ouest  
65 rue de Saint-Brieuc, CS 84215, 35042 Rennes cedex

soutenue le 16 juin 2008

devant le jury composé de :

Florence LE BER,	Ingénieure en Chef du GREF, ENGEES Strasbourg, rapporteur
Philippe LAGACHERIE,	Ingénieur de recherche, INRA Montpellier, rapporteur
Benoît GABRIELLE,	Professeur, AgroParisTech, examinateur
Delphine LEENHARDT,	Chargée de recherche, INRA Toulouse, examinatrice
Jean-Sébastien PIERRE,	Professeur, Université de Rennes 1, examinateur
Christian WALTER,	Professeur, Agrocampus Ouest, directeur de thèse
Patrick DURAND,	Directeur de recherche, INRA Rennes, co-directeur de thèse





*« La simplicité est la sophistication ultime. »*

Léonard de Vinci

*« There is a theory which states that if ever anyone discovers exactly what the Universe is for and why it is here, it will instantly disappear and be replaced by something even more bizarre and inexplicable.*

*There is another theory which states that this has already happened. »*

Douglas Adams (The Hitchhiker's Guide to the Galaxy)

Couverture : Monique Meyer, 2008, « Paysage virtuel », aquarelle.



## Abstract and key words

The environmental assessment of agrosystems is a critical issue for the design of sustainable agricultural practices and for the preservation or recovery of the ecological functions of agrosystems. However, this assessment is difficult to achieve through expert opinion and experimentation alone, due to the diversity of transformation and transfer processes of organic matter within agricultural landscapes, their interactions, and their dynamics. This project aimed at proposing a spatialized and dynamic environmental assessment method to integrate the effects of the physical environment, climate, and agricultural practices.

Our approach was structured by the development of a landscape modeling platform (**Qualscape**), which provided the conceptual framework for the different steps of this study:

- analysis of observed field crop sequences to identify the spatial and temporal drivers of crop transitions, because landcover plays a major role in the location of agricultural practices;
- the construction of a set of virtual agricultural landscapes based on the intensively farmed Naizin catchment (12 km<sup>2</sup>, Morbihan department, France) that differed in terms of climate, soil-type patterns, and crop management;
- the multi-target and multi-criteria assessment of these agrosystems by coupling and integrating the outputs of the nitrogen (N) fate model TNT2 and a soil phosphorus (P) budget model.

In addition, **Stochastree**, a crop-transition model based on stochastic decision trees, was developed. It maintained the spatial distribution of crops around the farmsteads while accounting for other constraints such as soil waterlogging and farm-production objectives. Environmental assessments of the virtual landscapes showed the relative influence of climate, soil-type patterns, and cropping systems on a set of agricultural, soil, and hydrological indicators. Although a moderate, less-intensive cropping system reduced soil P surpluses and improved N- and P-use efficiencies, climate series and soil-type patterns influenced the stream N concentration more heavily than cropping systems. Moreover, the magnitude of soil-climate interactions on stream N concentrations was equivalent under certain scenarios to a 15% reduction in the amount of total N applied. This shows that remediation practices aimed at achieving a specific environmental goal (such as not exceeding the 50 mg NO<sub>3</sub>.l<sup>-1</sup> threshold in streams) must take local landscape characteristics into account, and that transplanting them to another landscape may not attain the expected goal.

**Key words:** virtual landscape, environmental assessment, agrosystem, soil, climate, agricultural practice, fertilization, nutrient, nitrogen, nitrate, phosphorus, hydrology, TNT2, Qualscape, Stochastree, Rotomatrix

## Résumé et mots-clefs

L'évaluation environnementale des systèmes de production agricole est un enjeu majeur pour le développement de pratiques agricoles durables, la préservation ou la restauration du bon état écologique des ressources naturelles. Cependant, cette évaluation est compliquée à réaliser par expertise ou par expérimentation à cause de la diversité des processus de transformation et de transfert de la matière organique au sein des paysages agricoles, leurs interactions et leurs dynamiques. L'objectif de ce projet était de proposer une méthodologie d'évaluation environnementale qui soit spatialisée et dynamique afin d'intégrer les effets du milieu physique, du climat et des pratiques agricoles.

L'approche suivie dans ce projet est basée sur le développement d'une plateforme de modélisation des paysages agricoles (**Qualscape**), comme support conceptuel commun aux différentes étapes de ce travail :

- l'analyse de successions culturales parcellaires observées afin d'identifier les déterminants spatiaux et temporels des transitions de cultures ; l'occupation du sol spatialisant les pratiques agricoles ;
- à partir du bassin versant de Naizin (12 km<sup>2</sup>, Morbihan, France) caractérisé par des activités d'élevage bovin et porcin intensives, la construction d'un jeu de paysages agricoles virtuels différant par leur milieu physique, les forçages climatiques et les pratiques agricoles auxquels ils sont soumis ;
- la réalisation d'une évaluation environnementale multi-cibles et multi-critères de ces agrosystèmes en permettant le couplage et l'intégration des sorties entre le modèle agro-hydrologique TNT2 et un modèle de bilan de phosphore du sol.

**Stochastree**, un modèle de transition de cultures basé sur des arbres de décision stochastiques a été développé. Il a notamment permis de maintenir la distribution spatiale des cultures autour du corps d'exploitation, tout en respectant des contraintes d'hydromorphie des sols et d'objectif de production des exploitations. L'évaluation environnementale des paysages virtuels a quantifié l'influence du climat, du milieu et des systèmes de culture sur différents indicateurs agricoles, pédologiques et hydrologiques. Nous avons constaté que l'adoption d'un système de culture modéré (réduction d'intrants, intercultures hivernales) réduit significativement les flux excédentaires de phosphore vers les sols et accroît l'efficacité des apports d'azote et de phosphore. Cependant, cet effet s'atténue pour des indicateurs incorporant des processus de transformation et de transfert de la matière organique, comme la concentration en nitrate à l'exutoire. De plus, les effets des interactions sol-climat sur cette concentration sont du même ordre qu'une réduction de 15% des apports totaux d'azote. Ainsi, la définition de pratiques agricoles alternatives devant satisfaire un objectif environnemental précis (cf. le non-dépassement du seuil des 50 mg NO<sub>3</sub>.l<sup>-1</sup> dans les cours d'eau) ne peut se faire qu'en lien avec les caractéristiques du paysage et que leur transposition directe à un autre paysage peut ne pas atteindre l'effet escompté.

**Mots-clefs** : paysage virtuel, évaluation environnementale, agrosystème, sol, climat, pratique agricole, fertilisation, nutriment, azote, nitrate, phosphore, hydrologie, TNT2, Qualscape, Stochastree, Rotomatrix

## Remerciements

*« [Il] émergeait de nouveau, et reprenait, appuyé à gauche, sa veille terrible. Il ne savait plus combien de temps, combien d'efforts le délivreraient de ses liens sombres. Il doutait presque d'en être jamais délivré, car il jouait sa vie sur ce petit papier, sale et chiffonné, qu'il avait déplié et lu mille fois, pour bien nourrir son espérance »*

Saint-Exupéry (Vol de nuit)

*« ± Contente-toi de bosser comme un crétin, pas à pas. [1/4]  
± Tu me conseilles de bosser logiquement, au fond.  
± Oui, tu connais autre chose ?  
± Je ne connais que l'autre chose. »*

Fred Vargas (Dans les bois éternels)

Après une période post-soutenance relativement mouvementée, j'ai enfin trouvé une semaine pour rédiger mes remerciements. Il ne me faudra pas moins d'une semaine pour remercier tous les gens avec qui j'ai collaboré, ceux qui m'ont aidé, soutenu, inspiré ou changé les idées... Bref, ceux avec qui j'ai vécu ce projet de recherche.

Ça faisait quelques temps que je souhaitais faire une expérience de recherche et développement. Après un entretien, Christian Walter et Patrick Durand m'ont recruté sur ce beau projet combinant environnement et modélisation. Le co-financement de mon salaire par le centre INRA de Rennes et la Région Bretagne illustre les attentes scientifiques et territoriales du projet.

### Lundi

Lundi c'est le début de la semaine ; j'arrive à l'UMR SAS. Je suis accueilli chaleureusement par toute l'unité – dirigée par Chantal Gascuel – et tout le monde s'assure que je suis bien installé. La volonté générale de l'unité de donner de bonnes conditions de travail au personnel contractuel et de les intégrer dans les différents conseils de l'unité permet de leur faire prendre conscience des rôles qu'ils jouent dans la stratégie de recherche de l'unité. Et concernant les doctorants, d'en faire des chercheurs et des travailleurs plus responsables. J'espère que nos retours seront à la hauteur de leur ouverture d'esprit.

J'arrive dans l'unité avec Bruno Morel, recruté comme doctorant en même temps que moi. Il a l'air sympa. Tant mieux : nous allons partager le même bureau ! À peine nous commençons à traverser les couloirs ensemble que les gens nous confondent : et voilà, les Smorels sont nés !



## Mardi

Le projet prend forme, les pistes d'investigation fourmillent. Je suis conseillé par Christian et Patrick, ainsi que par le comité de pilotage suivant pour définir la bonne stratégie à mener : Laurence Hubert-Moy, Claudine Thenail, Benoît Gabrielle et Frédérick Garcia. Côté informatique, je découvre Linux avec Salim Combo, Eclipse et Java avec Jordy Salmon-Monviola. Hugues Boussard et Erwan Bocher me conseillent sur les aspects de modélisation spatiale du paysage. Le projet bénéficie de projets open-source qui nous font gagner beaucoup de temps :

- concernant Java : Eclipse<sup>1</sup>, GeoTools<sup>2</sup>, Java Topology Suite<sup>3</sup>, CSVReader<sup>4</sup>, JDom<sup>5</sup>, JMathPlot<sup>6</sup> ;
- mais aussi des logiciels de stockage, d'analyse et de synthèse de données : MySQL<sup>7</sup>, Weka<sup>8</sup>, R<sup>9</sup> et OpenOffice.org<sup>10</sup> ;

J'espère que la plateforme de simulation et d'analyse de paysages agricoles virtuels que je développe dans le cadre de ce projet pourra également être proposée pour utilisation et développement à la communauté open-source.

## Mercredi

Sur le volet agronomique de mon projet, j'ai reçu de bons coups de main de la part de Jean-Pierre Gillard, Safya Ménasseri, Françoise Vertès, Thierry Morvan, Hervé Squvidant, Fabien Ferchaud et Joël Belan. Du côté modélisation des exploitations agricoles, nous lançons une collaboration avec Christelle Raison, Philippe Faverdin et Xavier Chardon. J'apprécie beaucoup leur disponibilité et suis content de collaborer avec eux. J'espère que leurs projets se déroulent avec succès.

Un jour, un vieux copain alors en fin de doctorat, vient me taper sur l'épaule. Sébastien Couvreur me parle de l'ATIR, l'association des doctorants du centre INRA de Rennes. « Y a besoin de sang neuf » qu'il me dit. En lui disant que je ne suis pas sûr d'avoir beaucoup de temps à consacrer à la vie associative, je le suis quand même à l'assemblée générale de l'ATIR. Six mois de webmaster et un an de présidence plus tard, l'association a fait son bonhomme de chemin grâce à une équipe motivée et impliquée. Elle devient DocAIR afin de pouvoir intégrer tous les chercheurs contractuels INRA et Agrocampus.

Entre temps, les projets inter-associatifs menés au sein de Nicomaque<sup>11</sup> et de la CJC<sup>12</sup> m'ont donné la chance

---

1 <http://www.eclipse.org/>

2 <http://geotools.codehaus.org/>

3 <http://www.vividsolutions.com/JTS/JTSHome.htm>

4 <http://www.csvreader.com/>

5 <http://www.jdom.org/>

6 <http://jmathtools.sourceforge.net/doku.php?id=jmathplot:tutorial>

7 <http://www.mysql.com/>

8 <http://www.cs.waikato.ac.nz/~ml/weka/>

9 <http://www.r-project.org/>

10 <http://fr.openoffice.org/>

11 pôle rennais des associations de jeunes chercheurs: <http://www.nicomaque.org/>

12 confédération des jeunes chercheurs: <http://cjc.jeunes-chercheurs.org/>

de connaître et de collaborer avec tout un tas de gens sympathiques, en particulier Alban Cornillet, Maïwenn Corrigan, Morgane Gorria et Jérémie Lusseau lors de l'organisation du Forum Docteurs et Entreprises 2006. Le dynamisme et l'imagination palpables au sein de l'activité associative a été parfois compensée par la réactivité géologique de certains conseils académiques et leur écoute des représentants des jeunes chercheurs. Il y a eu cependant quelques personnes qui ont compris que nos idées n'étaient pas des revendications unilatérales mais des propositions destinées à faire mieux fonctionner le système de recherche dans son ensemble. Merci notamment à Catherine Réminiac, Daniel Boujard, Gérard Maise, et Grégoire Thomas de l'avoir compris et de nous avoir fait confiance. J'espère que cette cordialité entre générations continuera. À ceux qui croient que le système politique fonctionne et se régule « tout seul », voici la remarque d'Esther Joly qui m'a convaincu de m'investir aussi de cette façon pendant mon doctorat :

*« Les droits ne sont pas garantis par une autorité divine ; ils reposent sur la vitalité des institutions, des citoyens et des formations de pouvoir qui en soutiennent l'existence. Don't forget it. »*

## **Jeudi**

Valérie Viaud rejoint Christian et Patrick dans la supervision du projet qui bénéficie de nos échanges – et moi de leurs conseils et encouragements. Un projet de géostatistique se monte avec l'unité de mathématiques appliquées : je co-encadre avec Christian et David Causeur trois étudiants-ingénieurs motivés et persévérants dont le travail nous a ouvert de nouvelles perspectives de recherche. Merci Chloé Friguet, Cleo Tebby et Philippe Pacitto.

Au cours d'une discussion avec doudou Clément Ciani (guitariste tranquille devant l'éternel), nous lançons l'idée de monter un petit groupe de rock avec Christophe Le May, notre électrocardiogramme (plutôt « rock » qu'« électro » d'ailleurs) capable de rythmer notre adrénaline avec des troncs d'arbre. Baptiste Coquelle nous rejoint rapidement avec sa guitare mélodique et ses solos endiablés. Baptiste nous a sorti de l'anonymat ! Disons qu'il a commencé par nous donner un nom : Zikosmatic. Le mythe était né, restait à lui trouver une voix. De ce côté-là, nous avons eu la visite rapide d'un duo de choc – Émilie Yokoyama et Isabelle Le Guen – succédées par Ronan Le Cointe. Malheureusement pour nous, ils sont tous partis successivement pour un monde meilleur et nous avons dû chanter les morceaux nous-même. Un peu plus tard, histoire de venir la voir pour autre chose que lui demander son avis en statistiques, j'ai proposé à Maela Kloareg de venir jouer un peu de saxophone avec nous. Ça m'apprendra à faire l'intéressé : elle a réussi à me piquer la basse pour une des plus jolies lignes (ça lui va bien, tiens) qu'il nous a été donné de jouer, sous prétexte que je devais chanter et jouer de la guitare. C'est malin.

## **Vendredi**

Le temps se couvre, l'orage couve, mon soleil semble exilé si loin à l'Est que sans sa lumière, je me sens un peu faiblir. Je mets les bouchées doubles en me disant que cela le ramènera au zénith. La famille et les amis

sont heureusement là pour me soutenir et me changent les idées du mieux qu'ils peuvent (un quasi plein-temps pour Bruno, Alban et Sébastien). Ma Maman ensoleille mon esprit de ses pinceaux en dessinant le sympathique paysage agricole virtuel qui se trouve en couverture de ce rapport.

La journée est quasiment finie ; Bruno s'inquiète : le soleil s'est couché depuis longtemps et je n'ai pas l'air de vouloir en faire autant. Ça sent la nuit blanche et il n'aime pas ça : le lendemain matin, je ne suis plus forcément le même joyeux collègue (que j'espère être). Il sait que parfois je craque, et il se méfie quand le crack attack. En tout cas, il a une sacrée humanité à supporter ma tête et mes cernes de veilleur.

## **Samedi**

Cette dernière nuit courte a été bien plus constructive que d'autres grâce au soutien final sans faille et aux conseils de Christian pour la rédaction du rapport que vous êtes en train de lire. Grâce à lui, la dernière matinée se passe bien et quelques motivés dans l'unité m'assistent dans la préparation de la soutenance de mes travaux.

Christian et Patrick sont rejoints par Florence Le Ber, Philippe Lagacherie, Delphine Leenhardt, Benoît Gabrielle et Jean-Sébastien Pierre pour constituer mon jury de soutenance. J'ai eu beaucoup de plaisir à leur exposer mes travaux et à répondre à leurs questions. L'atmosphère « discussion entre collègues » que nous avons su installer a rendu les échanges particulièrement intéressants, c'est pourquoi je les remercie d'avoir joué leur rôle d'examineur de façon aussi stimulante.

La fête qui a suivi était géniale, le ton a été donné par les meilleures chansons de Zikosmatic (rappelez-vous... le mythe quelques paragraphes plus haut). Merci aux collègues, à la famille et aux amis d'être venus partager ce moment ensemble, ainsi qu'aux autres pour leurs encouragements à distance. Dans ces moments-là, tout compte.

C'est curieux, ce doctorat : il y a quelques mois, je ne rêvais que de tourner la page et de passer à autre chose. Aujourd'hui je me rends compte à quel point il est riche d'expériences humaines et qu'il ne s'agit pas de tourner irrévocablement la page, mais de passer aux chapitres suivants tout en gardant contact pour maintenir le fil de l'histoire.

## **Dimanche**

J'ai été un peu chamboulé par la semaine, merci à Euryale et Sébastien de m'avoir aidé à récupérer mes premières forces ; et à David et Malika d'avoir pris le relais en m'intégrant deux mois dans leur jeune famille aux côtés de Chloé. Hé hé hé... c'est moi qu'ai vu ses premiers pas, nananèèèère !

Lundi prochain, j'attaque une semaine dans une autre vie. Puisse-t-elle être aussi enrichissante et pleine de surprises dans les contacts qui m'attendent.

Mêmes souhaits de bonheur pour Émilie et sa famille.

# Table des matières

Abstract and key words.....	5
Résumé et mots-clefs.....	6
Remerciements.....	7
<b>Introduction générale.....</b>	<b>15</b>
Objectifs du projet.....	16
Enjeux.....	17
Structure du rapport.....	18
<b>Chapitre I : Modélisation intégrée et fonctionnelle des paysages et des pratiques agricoles.....</b>	<b>21</b>
Introduction.....	21
Paysage – modelage complexe et modélisation simplificatrice.....	23
Structuration du paysage agricole.....	24
Le milieu physique.....	24
Les composants agricoles du paysage.....	27
Conclusion partielle sur la structure des paysages agricoles.....	28
Fonctionnements hydrologique, pédologique et agricole du bassin versant.....	31
Les flux d'eau à l'échelle du paysage.....	31
Flux et transformation d'azote et de phosphore.....	33
Pratiques culturales et interactions eau-sol-plante.....	34
Conclusion partielle sur les dynamiques hydrologique, pédologique et agricole des bassins versants.....	37
Évaluation environnementale des agrosystèmes.....	39
Indicateurs et méthodologies d'évaluation.....	39
Modélisation et évaluation intégrée.....	42
Plateforme de modélisation des paysages agricoles.....	43
Apport de la POO à la modélisation intégrée du paysage.....	44
Conclusion, problématique et démarche adoptée.....	45
<b>Chapitre II : La plateforme Qualscape comme support de la modélisation des paysages agricoles.....</b>	<b>49</b>
Méthodologie générale – de la création de paysages virtuels agricoles à leur évaluation environnementale.....	49
Formalisme du paysage « occupation du sol » dans Qualscape.....	55
Le bassin versant de Naizin, support de la démarche.....	55
Une représentation hiérarchique du paysage agricole.....	57
Spécifications de la représentation du paysage pour le modèle Rotomatrix.....	57
Spécifications de la représentation du paysage pour le modèle Stochastree.....	58
Conclusion sur la modélisation de paysages agricoles dans Qualscape.....	59
Formalisme du paysage « pratiques culturales » dans Qualscape.....	60
Les sorties parcellaires du modèle TNT2.....	60
Paysage « pratiques culturales ».....	60
Modèle phosphore.....	62
Représentation du milieu physique.....	63
Spatialisation de l'épaisseur des sols.....	63
Spatialisation des types de sols.....	63
Structure verticale des sols dans TNT2.....	64
Sources de la plateforme Qualscape.....	65
Le projet Java.....	65
L'extension SpatialAnalysis.oxt.....	65
Les scripts R.....	66
<b>Chapitre III : Stochastree, un modèle de successions de cultures basé sur des arbres de décision stochastiques, intégrant des contraintes agronomiques.....</b>	<b>69</b>
Résumé français.....	69
Abstract.....	71
Introduction.....	72

Materials and methods.....	75
Presentation of the Naizin catchment.....	75
Agronomic constraints on the agricultural landuse of Naizin.....	76
Structure of the crop transition models.....	78
Comparative analysis of the simulations.....	82
Results.....	85
Driving factors of crop transition in Naizin.....	85
Structure of the crop transition models.....	85
Comparative analysis of the simulations.....	89
Discussion.....	93
Comparing observations and simulations.....	93
Validation and generalization of Stochastree simulations.....	94
Mutual benefits of decision trees and Markov chains approaches in LCM.....	95
Improving the construction of the decision trees.....	95
Conclusion.....	96
Acknowledgments.....	96
References.....	97
<b>Chapitre IV : Une approche multi-ressources et multi-critères à l'échelle du paysage, d'évaluation de l'impact environnemental des pratiques agricoles.....</b>	<b>103</b>
Résumé français.....	103
Abstract.....	105
Introduction.....	106
Materials and methods.....	108
Hydrological and nutrient fate models.....	108
Factorial simulation design.....	111
TNT2 simulations.....	115
Statistical analysis.....	115
Results.....	117
Virtual landscape construction.....	117
Model calibration.....	121
Scenario comparison and analysis.....	123
Discussion.....	129
Construction of the virtual landscapes.....	129
Multi-resource and multi-criteria analyses.....	130
Methodology improvements and perspectives.....	132
Conclusion.....	133
Acknowledgments.....	133
References.....	134
<b>Chapitre V : Conclusion générale.....</b>	<b>139</b>
Synthèse des résultats.....	139
Méthodologie générale.....	139
Déterminants spatiaux et temporels de l'évolution de l'occupation du sol.....	140
Évaluation environnementale des agrosystèmes.....	140
La plateforme Qualscape.....	143
Points stratégiques.....	144
Stochasticité et déterminisme.....	144
Validation de l'évaluation environnementale des paysages virtuels.....	144
Gestion de la matière organique.....	145
Perspectives – de l'évaluation environnementale à l'évaluation de la durabilité des agrosystèmes.....	147
Bibliographie générale.....	151

<b>Annexes I : Application d'une modélisation par paysage virtuel pour évaluer des stratégies d'échantillonnage d'agroécosystèmes.....</b>	<b>165</b>
Résumé français.....	165
Abstract.....	168
Introduction.....	169
Material and methods.....	171
Construction of the initial state of the virtual landscape.....	171
Virtual landscape evolution.....	175
Sampling strategies.....	179
Evaluation of sampling strategies.....	180
Results.....	183
Initial state of the soil P content distribution.....	183
Global evolution of soil P content.....	185
Strategy evaluation.....	187
Discussion and conclusion.....	190
References.....	192
<b>Annexes II : Matrices de probabilités de transition et arbres de décision stochastiques complétant les résultats du chapitre III et IV.....</b>	<b>197</b>
Matrices de transition de cultures pour les différents types d'exploitation du bassin versant de Naizin.....	197
Arbres de décision utilisés par Stochastree pour les différents types d'exploitation du bassin versant de Naizin.....	199
Arbre de décision utilisé pour simuler l'interculture hivernale.....	205
<b>Annexes III : Format et manipulation des arbres de décision.....</b>	<b>209</b>
Définition du format XML utilisé.....	209
Production et visualisation des arbres.....	209
Méthode principale de manipulation des arbres.....	210



## Introduction générale

Les paysages européens sont modelés par les activités rurales : 44% sont des terres agricoles et 33% sont des surfaces forestières dédiées en partie à la production sylvicole (European Commission, 1999). En tant qu'activité de production, l'agriculture s'appuie sur l'utilisation de ressources naturelles, exerçant de fait une pression sur l'environnement. En réponse aux besoins alimentaires d'après-guerre et à des exigences économiques de rentabilité et de compétitivité, l'agriculture s'est considérablement intensifiée durant ces 50 dernières années. Cette intensification de la production a augmenté la pression exercée sur l'environnement, s'accompagnant notamment d'un remodelage des paysages (destruction de haies, fusion de parcelles, drainage), d'une spécialisation des systèmes de production et d'un accroissement de la consommation d'intrants (fertilisants, pesticides).

La prise de conscience de la dégradation des ressources naturelles due aux impacts de l'agriculture en Europe progresse depuis les années 1980. La *Directive Nitrates* (adoptée en 1991) et la *Directive-Cadre sur l'Eau* (adoptée en 2000) ont établi un cadre pour une politique globale communautaire de reconquête de la qualité de l'eau, en se basant notamment sur la réduction des intrants. Cependant, si des liens ont été mis en évidence entre leur circulation accidentelle dans l'environnement et la dégradation des ressources naturelles, l'établissement d'une relation chiffrée entre les quantités appliquées et les impacts est loin d'être simple (Ruiz et al., 2002 ; Dubus et al., 2003 ; Basset-Mens et al., 2006 ; Deasy et al., 2008), car la structure du milieu physique et la répartition des pratiques agricoles dans l'espace modulent l'intensité et la vitesse de transferts des intrants au cours d'eau (Colin et al., 1999 ; Pauwels et al., 2001 ; Beaujouan et al., 2002 ; Molénat et al., 2002 ; Moussa et al., 2002). À ces sources d'incertitudes, se rajoute la variabilité climatique qui influence le développement des végétaux, leurs prélèvements de nutriments dans le sol et les flux d'eau au sein du territoire.

En fixant un seuil maximal de potabilité des eaux à 50 mg NO<sub>3</sub>.l<sup>-1</sup> et une recommandation de 25 mg NO<sub>3</sub>.l<sup>-1</sup> pour les nappes, la réglementation a encouragé le développement d'aménagements permettant de réduire la pression azotée exercée sur l'eau par des processus naturels (dénitrification hétérotrophe dans les sols de fond de vallée) ou artificiels (station de traitement des excédents d'azote). Cependant, si ces approches ont permis de réduire la pression azotée sur l'eau, elles ont conduit dans une certaine mesure à un transfert de pollution en la reportant partiellement sur l'atmosphère par des émissions non négligeables de gaz à effet de serre (Mosier et Kroeze, 2000). De surcroît, le raisonnement de la fertilisation par des effluents organiques sur la seule base de l'azote entraîne généralement des apports excédentaires de phosphore. Des travaux récents ont ainsi mis en évidence une augmentation des stocks de phosphore dans les sols des régions d'élevage intensif (Djodjic et al., 2005 ; Lemercier et al., 2006).

Ces faits montrent donc qu'il n'est pas suffisant d'évaluer l'impact environnemental des systèmes de production agricole en se basant sur une seule ressource (comme l'eau) ou sur un seul critère (comme la teneur en azote). Aussi, des législations départementales ou nationales ont renforcé l'exigence d'un bilan



azoté équilibré (entre les exportations par les récoltes et les fertilisations) par l'obligation d'un excédent moyen de phosphore à l'échelle de l'exploitation inférieur à des seuils pouvant aller de 9 kg P.ha<sup>-1</sup> (Aarts et *al.*, 2000) à 44 kg P par hectare de surface épanachable (Préfecture d'Ille-et-Vilaine, 2005).

La modification des pratiques agricoles constitue le levier majeur que nous pouvons saisir pour développer des politiques de reconquête de la qualité de l'eau et des sols ; et en contexte d'élevage intensif, la gestion de la matière organique est au cœur du problème. Cependant, du fait des interactions spatiales et temporelles complexes, l'évaluation de l'efficacité des programmes d'action est difficile à mener par avis d'experts, mais aussi par l'expérimentation. D'une part, le coût induit par le développement d'un réseau de mesure nécessaire au suivi des différents polluants risque d'être prohibitif, et d'autre part, cette évaluation est nécessaire *avant* la mise en place effective des programmes. Le recours à la modélisation semble donc indispensable pour la prise en compte simultanée des processus et des interactions évoqués précédemment, mais aussi pour la création et le test de scénarios alternatifs.

Les approches de modélisation dédiées à l'évaluation environnementale des systèmes agricoles (ou agrosystèmes) abordent une grande diversité de méthodologies, d'échelles spatiales et temporelles. Certaines sont fondées sur des scores qui leur offrent une certaine robustesse, mais une intégration limitée des spécificités du contexte de l'étude : des scores de durabilité agroécologique avec la méthode IDEA (Zahm et *al.*, 2005) ou les Phosphorus Indicator Tools (Heathwaite et *al.*, 2003) ; des scores ne concernant que la consommation de ressources naturelles, exprimée en « équivalent planète Terre » pour l'empreinte écologique (Wackernagel et *al.*, 1999). Le modèle IFSM (Integrated Farm System Model) développé par l'USDA (Rotz et *al.*, 2007) propose quant à lui une modélisation dynamique fine de la gestion agronomique d'un système de production animale, et permet d'évaluer ses performances à long terme ainsi que ses impacts environnementaux sous forme de facteurs d'émission. La diversité des approches brièvement exposée dans cette introduction suggère la nécessité d'élaborer un état des connaissances plus approfondi concernant la modélisation des paysages agricoles et de l'évaluation environnementales des agrosystèmes.

### **Objectifs du projet**

Par une approche de simulation de paysages agricoles virtuels, l'objectif de ce projet est d'évaluer les impacts environnementaux des agrosystèmes en considérant les interactions des effets du milieu physique, du climat et des pratiques agricoles. Les étapes envisagées sont :

- dans un premier temps : *(i)* à partir de données de terrain, extraire les déterminants spatiaux et temporels des transitions de cultures ; *(ii)* implémenter ces connaissances dans un modèle simulant l'évolution de l'occupation du sol d'un paysage agricole ; et enfin, *(iii)* évaluer les capacités du modèle à respecter ces déterminants ;

- dans un second temps : (i) construire un jeu de paysages agricoles virtuels différant quant à leur milieu physique et les forçages climatiques et les pratiques agricoles auxquelles ils sont soumis ; (ii) faire une évaluation comparative de leurs impacts environnementaux.

Les hypothèses majeures de ce projet sont les suivantes :

- l'état actuel des connaissances nous permet de simuler l'état initial d'un agrosystème et d'en simuler de façon réaliste l'évolution sous des forçages climatiques et des pratiques agricoles ;
- la spatialisation des pratiques agricoles sur un milieu physique, lui aussi structuré dans l'espace, influence significativement leurs impacts sur les ressources naturelles.

## **Enjeux**

Sur le plan opérationnel, ce projet a l'ambition de fournir des résultats et des acquis méthodologiques pour des applications réelles, notamment sur la modulation et l'adaptation des mesures de réduction des impacts des agrosystèmes sur l'environnement en fonction du contexte pédo-climatique.

Sur le plan scientifique, ce projet s'inscrit tout d'abord dans la problématique de l'élaboration de paysages agricoles virtuels, qui articulent structures spatiales pédologiques et processus d'allocation des cultures au parcellaire ; puis dans l'évaluation multi-critères de leurs impacts environnementaux, qui s'appuie sur une démarche de simulation simultanée du développement des cultures, des flux d'eau et le devenir des nutriments. Son originalité est donc dans son ambition d'intégration des connaissances et des outils dans une démarche cohérente allant de la construction du système modélisé à l'interprétation des résultats des simulations des différents modèles.

Sur le plan organisationnel, ce projet s'inscrit dans le thème transversal de l'unité mixte de recherche 1069 INRA-Agrocampus SAS (Sol, Agro- et hydrosystème, Spatialisation), appelé « MIMeauSA »<sup>13</sup>, qui porte sur la conception et l'utilisation de modèles dynamiques spatialisés intégrant des processus biophysiques et soumis à des forçages anthropiques. À ce titre, il a vocation à stimuler des collaborations fortes entre les deux équipes de l'unité : FIELDS (Filière des effluents d'Élevage : évaluation environnementale des systèmes agricoles et qualité des Sols) et DEBA (Dynamique de l'Eau et des éléments associés dans les Bassins versants Agricoles).

---

13 Modélisation Intégrée Meso-échelle des impacts de l'agriculture sur la qualité de l'Eau, du Sol et de l'Air

## ***Structure du rapport***

Ce rapport s'articule autour de quatre parties :

- la première est consacrée au bilan des connaissances concernant la modélisation des paysages agricoles et l'évaluation des impacts environnementaux des agrosystèmes. Elle servira à finaliser le cadre général utilisé pour construire et évaluer les paysages virtuels agricoles ;
- la seconde décrit l'élaboration de *Qualscape*, plateforme de construction de paysages agricoles virtuels et d'analyse de leurs impacts environnementaux. Les deux parties suivantes présentent les résultats obtenus grâce à ces développements méthodologiques ;
- la troisième partie présente les résultats concernant l'apprentissage et la modélisation des transitions de cultures au sein d'un paysage agricole ;
- la quatrième partie traite de la construction et l'évaluation environnementale des paysages agricoles virtuels.

Nous concluons sur les principaux résultats de ce projet, les acquis méthodologiques et les perspectives ouvertes.

Les parties 3 et 4 sont présentées sous la forme d'articles soumis. Un troisième article propose une utilisation d'un paysage virtuel agricole comme support pour évaluer des réseaux de suivi de la qualité des sols. Cet article a été mis en annexes car son objectif nécessitait une conception du paysage virtuel qui ne correspondait que partiellement aux besoins et à la démarche générale du projet développé dans ce rapport.

## **Chapitre I**

Modélisation intégrée et fonctionnelle  
des paysages et des pratiques agricoles



# **Chapitre I : Modélisation intégrée et fonctionnelle des paysages et des pratiques agricoles**

## **Introduction**

Ce chapitre a pour objectif de faire une synthèse des connaissances et des concepts mobilisés dans la simulation de paysages agricoles, la modélisation de leur fonctionnement dans un contexte d'évaluation de leurs impacts environnementaux. Nous aborderons successivement :

- la description du milieu physique et des composants agricoles du paysage ;
- les flux d'eau et des nutriments associés au sein du paysage ;
- la modélisation et l'évaluation intégrée des agrosystèmes.

Les deux premières parties présenteront au fur et à mesure des approches spécifiques de modélisation, alors que la dernière partie s'attachera à identifier des points critiques dans l'évaluation intégrée des agrosystèmes et à proposer un cadre général pour tenter d'y répondre.

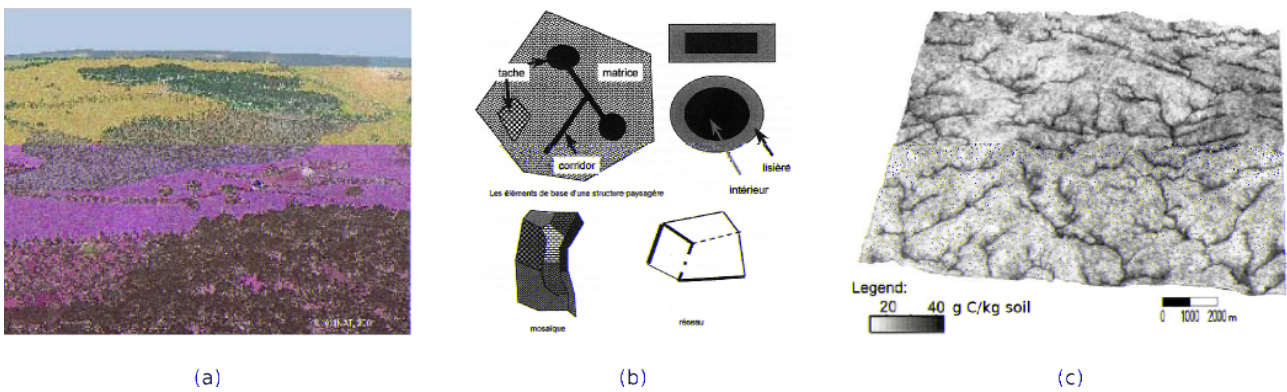


Fig. I.1: Trois regards disciplinaires différents sur le paysage :

- (a) modélisation du rendu visuel d'un projet d'aménagement forestier (Alinat, 2003) ;
- (b) les catégories d'éléments du paysage en écologie (Burel et Baudry, 1998) : la matrice est l'élément dominant et englobant ; les taches sont des zones homogènes (bosquets, parcelles, etc.), chacune délimitée par une lisière entourant un milieu intérieur ; les corridors sont des éléments linéaires (haies, ripisylve) qui peuvent être organisés en réseau ; l'ensemble des taches constitue une mosaïque ;
- (c) bloc-diagramme représentant une distribution spatiale de la teneur en carbone du sol, drapée sur le relief du paysage (Walter et *al.*, 2003).

# Paysage – modelage complexe et modélisation simplificatrice

Le concept de paysage est utilisé dans de nombreuses disciplines (aménagement, géographie, écologie, etc.) ; sa définition et sa représentation varient en fonction de l'objectif de ceux qui l'étudient (Fig. I.1). Dans son acceptation courante, il renvoie à la notion de portion de territoire vue et saisie par le regard de l'observateur (Fig. I.1.a). Nous retiendrons ici une définition plus fonctionnelle, telle que celle proposée par l'écologie du paysage, où il correspond à un système écologique au sens large qui est à la fois :

- hétérogène : il est constitué par un assemblage d'éléments dont l'organisation résulte de l'interaction entre le milieu physique, le climat et les activités humaines, d'où un modelage complexe ;
- dynamique : ses éléments évoluent et interagissent à plusieurs échelles spatiales et temporelles (Forman et Turner, 1981) ;
- le support de mouvements de matière, d'énergie et d'organismes vivants (Burel et Baudry, 2003).

Dans cette acceptation, la notion de paysage renvoie également à un niveau d'organisation pertinent permettant d'étudier simultanément son organisation spatio-temporelle et les processus qui s'y déroulent. La taille d'un paysage est généralement comprise entre l'étendue locale (microsite, parcelle, écosystème) et les étendues sub-régionales ou nationales. Mais son extension spatiale et la façon dont il est caractérisé dépendent des processus et des fonctions étudiées. Par exemple, les écologues décriront plutôt les habitats des espèces vivantes qu'ils étudient ainsi que la connectivité spatiale et temporelle des espaces de vie ou de circulation, que ce soit pour des approches relatives à la biodiversité ou à la gestion de ressources naturelles (Forman et Turner, 1981 ; Burel et *al.*, 1998 ; Gustafson, 1998). Aussi, l'agencement et la co-existence des éléments au sein du paysage offrent un degré d'analyse ou de représentation supplémentaire par rapport à l'énumération de ces éléments (Forman, 1995 ; Wielemaker et *al.*, 2001). En pédologie, Walter et *al.* (2003) ont mis en rapport des caractéristiques de surface du paysage (le parcellaire agricole), des pratiques agricoles avec un indicateur dérivé du relief afin de simuler l'évolution des stocks de carbone organique du sol. Le paysage correspond donc bien à une représentation d'un espace géographique, dont l'organisation et le fonctionnement ont été simplifiés pour répondre à un objectif précis.

Ce dernier point amène une remarque commune aux activités de modélisation au sens large (qu'il s'agisse de représentation ou de fonctionnement), que je souhaite présenter dès à présent car elle est restée en filigrane tout au long du développement méthodologique et de la réalisation du travail exposé dans ce rapport. Il s'agit du concept d'« unknowability »<sup>14</sup> du paysage, problème fondamental relevé par Beven (2002) au sujet de l'étude des systèmes hydrologiques, et généralisable à tous les systèmes faisant intervenir des phénomènes se produisant sous terre. En effet, malgré le développement des techniques d'investigation du paysage, *in situ* (prélèvement, sonde, tomographie) ou à distance (télétection), notre capacité à observer et à décrire la

---

14 « unknowability » du paysage : néologisme qu'on pourrait traduire approximativement par « inconnaissabilité »



surface et le sous-sol est limitée dans l'espace et dans le temps. Ainsi, qu'il s'agisse de tester des hypothèses de description ou de fonctionnement, la modélisation du système nécessite un compromis entre une description détaillée et notre capacité à la renseigner. Il en va aussi de l'aptitude du modèle à être utilisé, testé ou validé dans d'autres contextes *a priori* autant « inconnus » que le premier. En pratique, il en résulte souvent que différents compromis de modélisation d'un système environnemental peuvent délivrer des résultats similaires en terme de vraisemblance, sans qu'un optimum puisse être atteint. Ce principe d'équifinalité<sup>15</sup> (Beven et Freer, 2001) suggère même de rejeter l'idée d'un modèle optimal, ou d'un jeu de paramètres optimal, lorsque le système étudié est décrit par un jeu de données restreint. De même, pour chacun des thèmes abordés dans la suite de cette synthèse, le panel des alternatives de modélisation présenté a été concentré, par pragmatisme, aux approches pouvant pertinemment répondre à notre objectif de recherche.

Dans le cadre de ce travail, nous nous intéressons aux paysages agricoles, c'est-à-dire ceux dont la surface est majoritairement dédiée aux cultures, aux prairies, ou à l'élevage. Les paysages européens sont modélisés par les activités agricoles : 77% de l'espace est occupé par des paysages agricoles, dont 44% sont des terres agricoles et 33% des surfaces forestières gérées en partie par des agriculteurs (European Commission, 1999). De nombreuses études, qui vont du sondage sur le terrain à la comparaison de données de télédétection sur plusieurs pays, ont permis de dégager un certain nombre de concepts permettant de décrire ces paysages agricoles, de décrire les flux, de modéliser leur fonctionnement de façon intégrée afin d'en évaluer les impacts environnementaux. C'est ce fil conducteur que nous allons suivre.

## **Structuration du paysage agricole**

### ***Le milieu physique***

Le sol, support des cultures, constitue une interface complexe et encore relativement méconnue entre l'atmosphère, la lithosphère, l'hydrosphère et la biosphère. Leur formation fait interagir des processus d'altération de la roche, de remaniements chimiques (liés en particulier aux flux d'eau), mécaniques (compression, glissement de terrain) et biologiques (bioturbation) complexes et intervenant à différentes échelles spatiales et temporelles, qu'il ne s'agit pas de présenter ici.

### **Variabilité verticale des sols**

Les premiers travaux initiateurs de la pédologie ont d'abord consisté à décrire localement des profils verticaux de sols, à en comprendre les mécanismes de formation et d'évolution (Dokuchaev, 1883). Les profils pédologiques sont généralement représentés selon une approche compartimentale, par une succession d'horizons<sup>16</sup>, chacun caractérisé par leur composition, mais aussi par leurs propriétés structurales (compacité, porosité, conductivité hydraulique, etc.). Les pédologues ont conçu des référentiels pédologiques nationaux (Baize et Girard, 1995) et internationaux (IUSS Working Group WRB, 2007) classant les sols en fonction de

---

<sup>15</sup> le terme utilisé par Beven est « equifinality », mais on trouve le néologisme « équifinalité » dans de nombreuses communications

<sup>16</sup> horizon d'un sol : couche horizontale de sol homogène dans sa texture ou sa structure (Baize et Girard, 1995)

successions caractéristiques et qui permettent, après cartographie, d'étudier la distribution spatiale des types de sols. Par « distribution spatiale » ou « spatialisation », on désigne généralement la variabilité d'une propriété sur le plan horizontal, sous forme de carte par exemple.

### Variabilité horizontale des sols

L'étude de la variabilité spatiale des sols a récemment bénéficié de l'essor des Systèmes d'Information Géographique (SIG) et de la télédétection concernant la représentation des données, et de l'analyse statistique et géostatistique pour leur traitement (McBratney et al., 2003 ; Walter et al., 2005 ; Pennock et Veldkamp, 2006 ; Sumfleth et Duttmann, 2008). En fonction de la nature de la variable considérée, deux modes de représentation sont principalement utilisés :

- représentation *vectorielle* : l'utilisation de points, de lignes et de polygones permet de décrire différents éléments du paysage (Fig. I.1.b). Les polygones sont couramment utilisés pour dessiner des zones homogènes, comme la surface occupée par un même type de sol ou caractérisée par la même classe d'une variable catégorielle (occupation du sol, identité d'une parcelle) ;
- représentation *raster* : la représentation de variables continues nécessite une discrétisation dans l'espace réalisée généralement sous la forme d'une grille régulière à maille carrée. La valeur du pixel peut être représentée soit sous la forme d'un niveau de gris (comme la teneur en carbone du sol en nuances de gris sur la Fig. I.1.c), soit par une vue 3D (comme l'altitude sur la Fig. I.1.c).

La cartographie d'une propriété du sol se fait généralement à partir d'un nombre réduit d'observations (par échantillonnage de l'espace) par exemple, qu'il est nécessaire d'interpoler sur l'étendue de la carte. Parmi les nombreuses méthodes permettant d'interpoler une variable numérique, la géostatistique exploite son auto-corrélation spatiale, qui consiste à supposer que la différence de valeur entre deux points est fonction notamment de la distance qui les sépare. Un variogramme expérimental, représentant la variance de la variable d'intérêt en fonction de la distance séparant deux points peut être construit grâce à la formule suivante (Cressie, 1991), dans des conditions d'isotropie<sup>17</sup> et de stationnarité<sup>18</sup> :

$$\gamma(h) = \frac{1}{|N(h)|} \sum_{(i,j) \in N(h)} (Z_i - Z_j)^2 \quad (1)$$

où  $\gamma(h)$  est la variance moyenne de la variable  $Z$ , mesurée empiriquement entre des points séparés par la distance  $h$  ;  $N(h)$  est l'ensemble des paires de points  $i$  et  $j$ , séparés par la distance  $h$  ;  $|N(h)|$  est le nombre de paires de points dans  $N(h)$ .

17 isotropie : hypothèse forte supposant que la variance dépend uniquement de la distance, indépendamment de la direction

18 stationnarité : 2<sup>e</sup> hypothèse forte, supposant que la relation établie par le variogramme ne varie pas dans l'espace

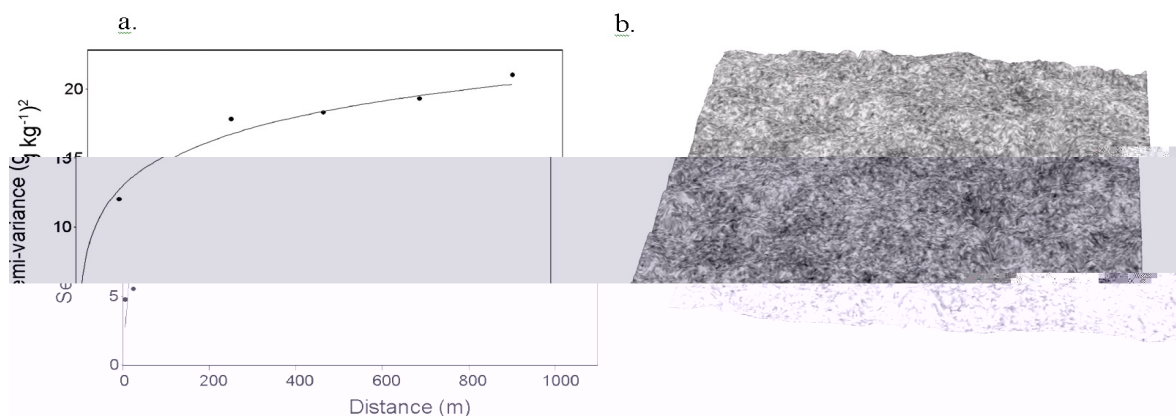


Fig. I.2: (a) Variogramme expérimental de la teneur en carbone du sol calculé à partir de 360 échantillons, (b) résultats d'une simulation conditionnelle faite à partir des valeurs des 360 échantillons et du variogramme empirique (Walter *et al.*, 2003).

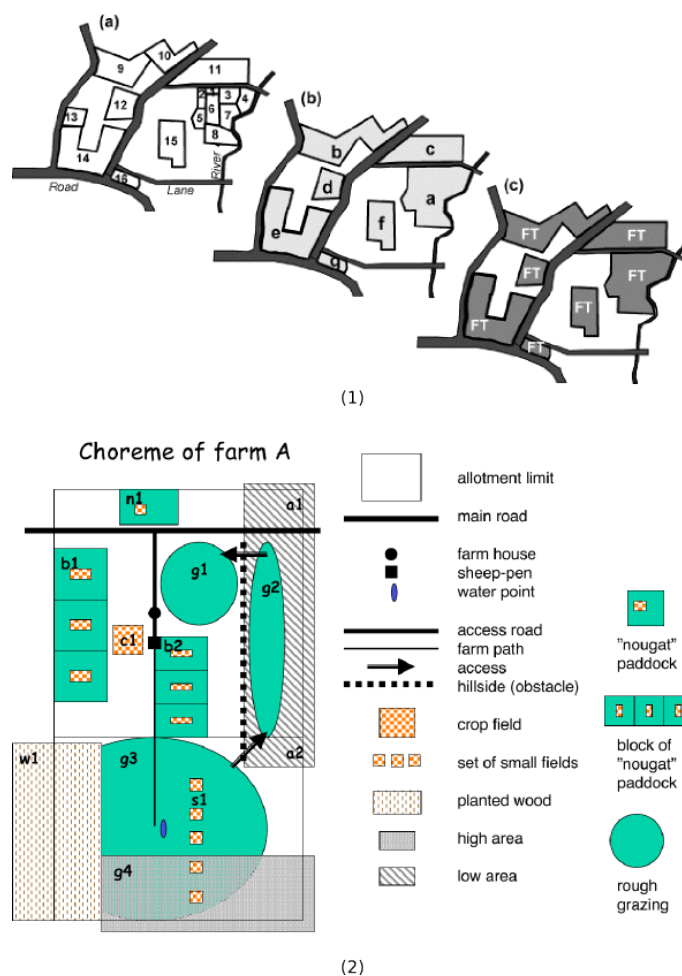


Fig. I.3: Deux démarches de représentation fonctionnelle des unités de gestion du territoire d'une exploitation : (1) hiérarchisation des unités paysagères agricoles par agrégation spatiale (Thenail et Baudry, 2004) : (a) les parcelles ; (b) les îlots ; (c) l'exploitation agricole (FT = farm territory) ; (2) représentation chorématique des relations topologiques et fonctionnelles entre les éléments d'une exploitation agricole (Le Ber *et al.*, 2003) : (a) bâtiment, (b, c, n) parcelles, (g) pâtures, (w) forêt. Les chorèmes de liaison informant sur la connectivité entre ces éléments et leur rapport avec la topographie générale.

Ces variogrammes expérimentaux peuvent être approchés par des fonctions mathématiques simples (Cressie, 1991) et utilisables par des algorithmes d'interpolation ou de simulation conditionnelle afin de produire des cartes (Fig. I.2) modélisant des composantes de paysages (Walter et al., 2003). Des approches similaires ont été développées pour des variables catégorielles à partir de lois de voisinages (Lagacherie, 1992), ou à partir de transiogrammes, structures similaires aux variogrammes donnant, en fonction de la distance à un point dont la classe est connue, les probabilités qu'un point non classé appartienne aux différentes classes possibles (Zhang et Li, 2008). McBratney et al. (2003) ont montré que la diversité des méthodes utilisables pour créer ou interpoler des cartes pédologiques couvre désormais un large spectre de techniques (modèles linéaires généralisés, réseau de neurones, logique floue, etc.) et que les défis, aujourd'hui, sont de s'approprier les théories sous-jacentes, d'avoir des données de qualité pour les mettre en œuvre et de l'expérience pour éviter un sur-apprentissage des modèles aux données.

### **Les composants agricoles du paysage**

Les structures agricoles considérées en agronomie sont généralement présentées par niveau d'agrégation spatiale (Thenail et Baudry, 1994 ; Deffontaines et al., 1995 ; Thenail et Baudry, 2004) (Fig. I.3.1) :

- la parcelle : c'est l'unité de gestion élémentaire du territoire agricole. Les interventions culturales (travail du sol, semis, fertilisation, traitement phytosanitaire, récolte) y sont réalisées de façon homogène sur toute sa surface, aux modulations près en agriculture de précision ;
- l'îlot de parcelles : les parcelles d'un îlot peuvent être associées pour accueillir la même culture et sont souvent gérées de façon synchrone ;
- le territoire de l'exploitation : il est géré de façon cohérente pour répondre à des besoins et des objectifs de production ;
- le siège d'exploitation : il regroupe habituellement l'essentiel du matériel agricole et les éventuels bâtiments d'élevage. Pour les exploitations laitières avec pâturage, l'accessibilité des parcelles à partir du siège d'exploitation peut jouer un rôle déterminant dans l'allocation des cultures et des prairies sur le parcellaire notamment dans la régulation de la charge de travail ;
- les éléments intersticiels : les haies, par exemple, constituent un réseau de corridors ou de barrières pour les populations d'insectes (Hunter, 2002) ainsi qu'une zone tampon hydrologique et géochimique (Viaud et al., 2004). Les fossés de route ou de drainage modifient la circulation des eaux de surface (Moussa et al., 2002).

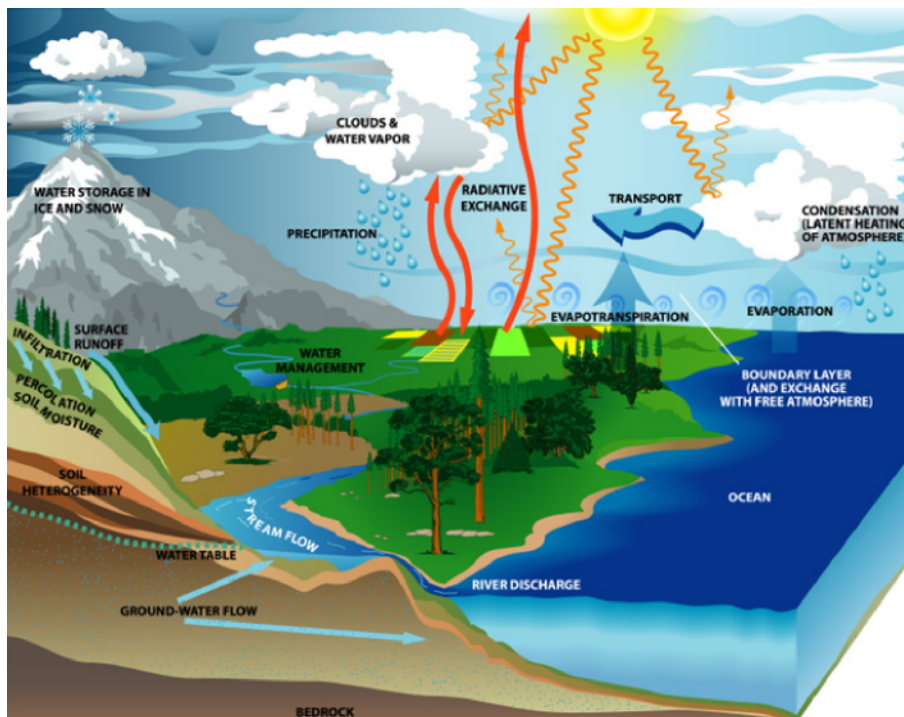
L'utilisation de chorèmes (Brunet, 1986) permet d'approfondir la description des liens topologiques et fonctionnels entre les différents éléments d'une exploitation agricole (Fig. I.3.2), ainsi qu'avec des éléments extérieurs à l'exploitation comme les accidents de relief, les réseaux hydrographique et routier, et des sites touristiques (Le Ber et al., 2003 ; Houdart et al., 2004). La prise en compte simultanée de la typologie des exploitations et des espaces fait de la modélisation graphique par chorèmes une approche originale de l'étude

des interactions entre les exploitations et les ressources environnementales, au sens large (ressources en eau, sols, patrimoines urbain et touristique).

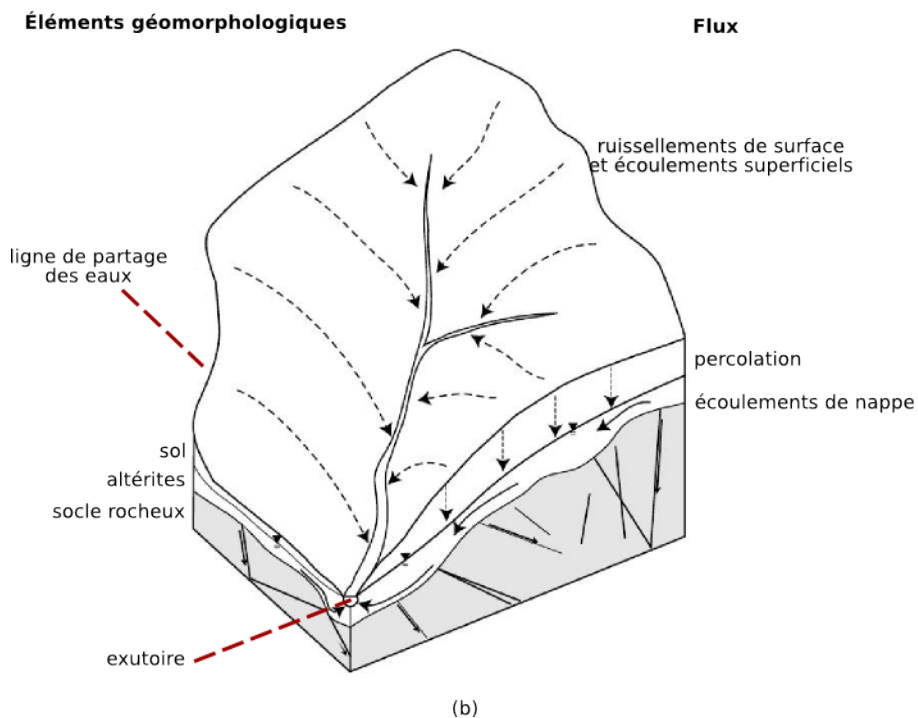
### ***Conclusion partielle sur la structure des paysages agricoles***

Les différentes disciplines qui se sont intéressées aux paysages agricoles nous offrent un panel très large de modes de représentation, d'analyse et de simulation de la structure du paysage. Ces modes de représentation du paysage ont évolué d'une approche cartographique et visuelle vers une approche analytique (McBratney et *al.*, 2003) et fonctionnelle (Le Ber et *al.*, 2003), intégrant les interactions spatiales entre les éléments du paysage. Cette partie a permis de construire une grille de lecture du paysage et de poser les bases pour l'étude de son fonctionnement dynamique, notamment en identifiant les unités spatiales de support aux flux d'eau et d'implémentation des pratiques agricoles.





(a)



(b)

Fig. I.4: Schémas du cycle hydrologique et d'un bassin versant :  
 (a) flux d'eau aériens, de surface et souterrains (US GCRP, 2003) ;

(b) représentation d'un bassin versant, délimité par la ligne de partage des eaux qui sont acheminées jusqu'à son exutoire (adapté de McGuire et McDonnel, 2006).

# Fonctionnements hydrologique, pédologique et agricole du bassin versant

## ***Les flux d'eau à l'échelle du paysage***

Un grand nombre d'étapes du cycle hydrologique se déroulent dans le paysage agricole (Fig. I.4.a) et les flux d'eau qui y ont lieu sont les moteurs de transferts conséquents d'azote (N) et de phosphore (P) (Carpenter et al., 1998 ; Sharpley et al., 2000 ; Hart et al., 2004). Les précipitations peuvent générer du ruissellement de surface lorsque le sol est complètement saturé (*ruissellement sur zone contributive*) (Kirkby, 1978), ou lorsque l'intensité de la pluie est supérieure à la capacité d'infiltration du sol (*ruissellement hortonien*) (Horton, 1933). Ces ruissellements s'écoulent en surface par gravité, avec d'éventuelles infiltrations et exfiltrations, jusqu'à rejoindre le cours d'eau. Les éléments du paysage agricole susceptibles de perturber significativement les ruissellements de surface sont les haies, qui peuvent détourner le parcours des ruissellements ou forcer leur infiltration par effet de barrage (Mérot et al., 1995), et les fossés de parcelles qui modifient leur cheminement (Moussa et al., 2002).

Quant aux eaux infiltrées, elles s'écoulent principalement verticalement en milieu non saturé jusqu'à rencontrer une couche imperméable du sous-sol. Les nappes d'eau formées par accumulation s'écoulent ensuite essentiellement latéralement jusqu'en fond de vallée où elles peuvent rejoindre le cours d'eau. La surface contribuant, grâce à ces écoulements de surface et de sub-surface, aux flux d'eau débités à un exutoire définit son bassin versant (Fig. I.4.b).

Dresser la liste des modèles hydrologiques développés à ce jour serait fastidieuse, Beven y a lui-même renoncé avec humour (2001), constatant la « pléthore » de modèles pluie-débit. Par contre, on retient que :

- la majorité des approches de modélisation hydrologique suit une démarche de compartimentation de l'espace dont la résolution spatiale horizontale peut aller du pixel de quelques mètres carrés au sous-bassin versant ;
- la topographie est une information essentielle, utilisée ne serait-ce que dans la délimitation des contours des bassins versants, mais aussi pour simuler la propagation des écoulements de surface, voire de sub-surface ;
- lorsque les modèles simulent les flux d'eau de sub-surface, une compartimentation verticale s'ajoute au maillage horizontal, et les écoulements sont modélisés par des vidanges verticales et latérales de ces compartiments les uns dans les autres ;
- le choix d'un maillage 3D plus ou moins fin est généralement la conséquence du souhait de simuler l'évolution temporelle du système hydrologique (notamment la saturation en eau de ces compartiments), afin de reproduire des variations fines de débit ou de hauteur de nappe dans le temps ;



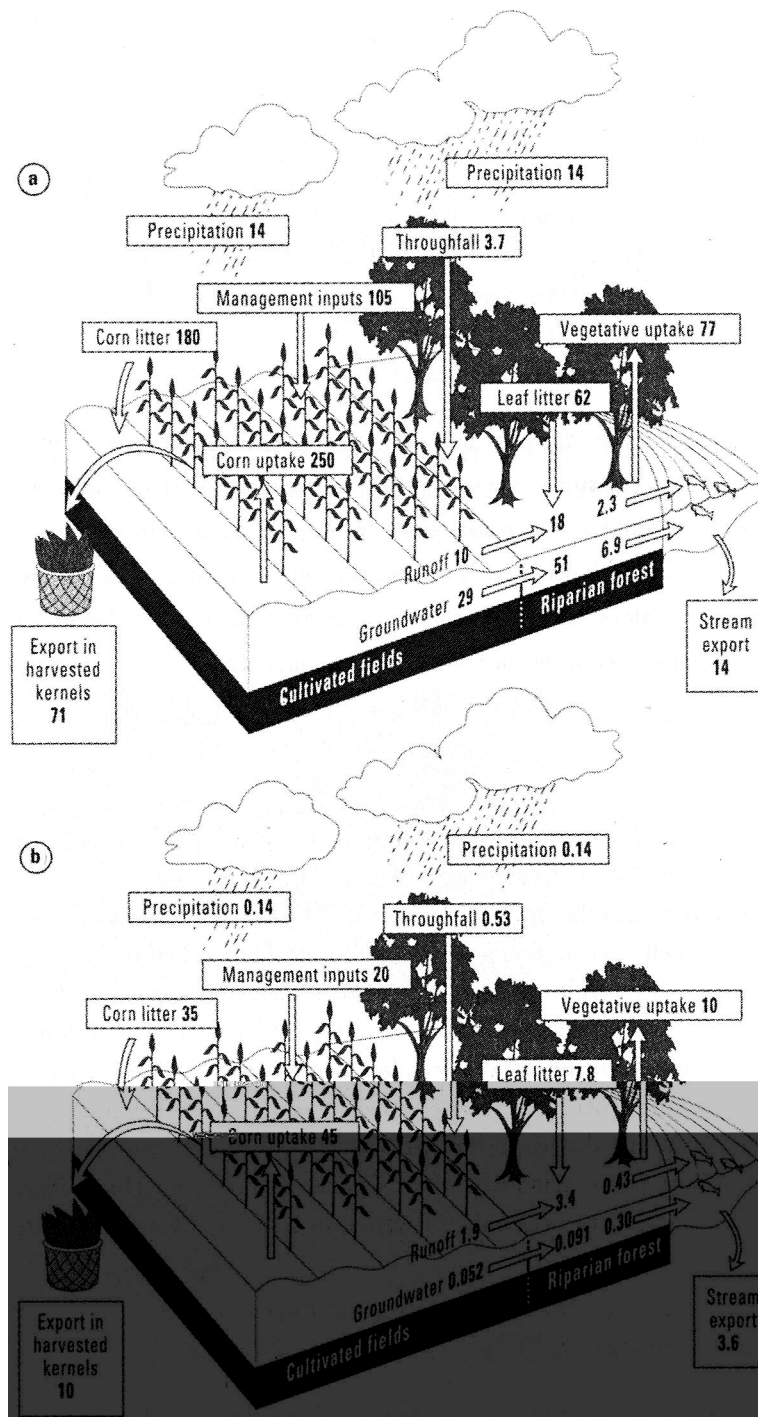


Fig. I.5: Flux d'azote (a) et de phosphore (b) mesurés au cours d'un cycle annuel culturel et hydrologique sur un sous-bassin versant de 3 ha constitué d'une parcelle de maïs séparé du cours d'eau par une forêt riparienne. Les flux sont exprimés en kilogramme par hectare de parcelle ou de forêt riparienne (adapté de Peterjohn et Correl, 1984).

- d'une façon générale, les choix de modélisation doivent être déterminés par une approche pragmatique reposant sur l'objectif de modélisation (test d'hypothèse de fonctionnement d'un bassin versant précis, construction d'un modèle robuste pour faire de la prédiction, construction d'un modèle générique comme outil d'aide à la décision) et la qualité des données utilisées pour conceptualiser, construire et tester le modèle.

Concernant ce dernier point, il est donc nécessaire d'appréhender les flux des éléments associés aux mouvements hydrologiques, ainsi que l'influence des pratiques agricoles sur ces phénomènes, avant de pouvoir envisager une modélisation intégrée d'un paysage agricole.

### **Flux et transformation d'azote et de phosphore**

La figure I.5 présente les flux d'azote (N) et de phosphore (P) entrant et sortant d'un territoire agricole. Les cycles de transformation du N et du P sont complexes et font intervenir de nombreuses formes moléculaires<sup>19</sup>, mais ces nutriments sont transportés essentiellement sous leur forme nitrate ( $\text{NO}_3^-$ ) et ammoniac ( $\text{NH}_4^+$ ) et de phosphate ( $\text{PO}_4^{3-}$ ). C'est également sous ces formes que les plantes les assimilent, et donc sous des formes très proches qu'on les synthétise pour faire des engrais minéraux (Carpenter et *al.*, 1998 ; Sharpley et *al.*, 2000 ; Hart et *al.*, 2004). Les durées de transferts de nitrate par les nappes jusqu'au cours d'eau peuvent dépasser plusieurs années pour des petits bassins versants ( $5 \text{ km}^2$ ) (Pauwels et *al.*, 2001 ; Molénat et Gascuel-Oudou, 2002), ce qui fait des nappes un réservoir de nitrate très important dont les fluctuations de hauteur peuvent être le facteur principal de contrôle de la concentration en nitrate à l'exutoire (Molénat et *al.*, 2008). Par contre, l'ion phosphate est mieux retenu par le sol du fait d'un plus grand nombre de charges ioniques, et de ce fait, les pertes de P se font majoritairement sous forme particulière lorsque des particules de sol sont arrachées par les ruissellement de surface (Hart et *al.*, 2004). Ces pertes par ruissellement de surface représentent moins de 10% de la quantité généralement apportée par fertilisation, mais ces faibles quantités suffisent à provoquer des phénomènes d'eutrophisation (Brazier et *al.*, 2005 ; Hart et *al.*, 2004). Étant donné les liens forts entre les flux d'eau et de N et de P, le bassin versant constitue une unité de gestion de la qualité de l'eau privilégiée, d'autant plus que 70 à 90 % des eaux des grands fleuves proviennent de bassin versants d'ordre<sup>20</sup> 1 à 3, c'est-à-dire en pratique, d'une surface d'une à quelques dizaines de kilomètres carrés (Kirkby, 1978 ; Vought et *al.*, 1994). Il s'agit donc de surfaces suffisamment petites sur lesquelles les collectivités territoriales peuvent développer des programmes de suivi et de reconquête de la qualité de l'eau, comme le programme Bretagne Eau Pure par exemple.

Le dépôt atmosphérique peut être une source non négligeable d'azote, estimée entre 5 et 50  $\text{kg N} \cdot \text{ha}^{-1} \cdot \text{an}^{-1}$  en contexte agricole (Berthe et *al.*, 2005). L'intensité des dépôts est en partie fonction des phénomènes de volatilisation qui peuvent se produire, pour les sources agricoles, dans les bâtiments d'élevage ainsi qu'au pâturage (Portejoie et *al.*, 2002). De plus, les conditions de stockage des effluents d'élevage en bâtiment

<sup>19</sup> des présentations détaillées de ces cycles ont été réalisées par Haynes (1986) pour N, Fardeau et Conesa (1994) pour P

<sup>20</sup> dans l'ordination des bras de rivière établie par Strahler (1952), tout bras « source » (qui n'a pas d'affluent) reçoit l'ordre 1 ; lors d'une confluence de deux bras d'ordre  $i$ , le bras résultant reçoit l'ordre  $i+1$  ; lors d'une confluence de bras d'ordre différents, le bras résultant reçoit l'ordre le plus grand des bras affluents

influencent les phénomènes de transformation de la matière organique, qui pourra avoir des propriétés variables (humidité du produit, teneur des éléments, forme moléculaire) au moment de son épandage sur les parcelles cultivées (Portejoie et *al.*, 2002). Enfin, les sols sont également le siège de deux processus de transformation des nutriments. La minéralisation de la matière organique (résidus de récoltes, débris racinaires) peut libérer sous forme lessivable plus de 150 kg N.ha<sup>-1</sup>.an<sup>-1</sup>, mais ces valeurs varient en fonction du type de sol (Nicolardot et *al.*, 1996). La dénitrification, processus qui réduit le nitrate sous forme gazeuse (NO, N<sub>2</sub>O et N<sub>2</sub>), peut atteindre des taux de 30 kg N.ha<sup>-1</sup>.jour<sup>-1</sup> dans les sols hydromorphes de vallées (Fig. I.5), où la saturation en eau impose les conditions d'anoxie nécessaires au déclenchement de la respiration anaérobie des bactéries dénitrifiantes (Durand et *al.*, 1998). Les processus de dénitrification ont été observés majoritairement en surface (Hénault et *al.*, 2000), jusqu'à 40 cm de profondeur et avec une variabilité spatiale importante de leur intensité (Oehler et *al.*, 2007).

## **Pratiques culturales et interactions eau-sol-plante**

### **Distribution séquentielle et spatiale des pratiques culturales**

La chronique des opérations faites sur une parcelle pour la conduite d'une culture constitue un itinéraire technique. Les opérations culturales influençant potentiellement les flux d'eau et d'éléments associés sont :

- le travail du sol : il rétablit l'état fragmentaire et poreux du sol et diminue ainsi généralement les phénomènes de ruissellement de surface et les risques d'érosion (Boiffin, 1984) ;
- les fertilisations : « minérales » ou « organiques » selon qu'il s'agisse d'engrais de synthèse ou d'effluents, les fertilisations visent à apporter les nutriments nécessaires au développement des cultures. Les fertilisations enrichissent donc la surface du sol en éléments susceptibles d'être emportés par ruissellement ou lessivage au prochain événement pluvieux (Hart et *al.*, 2004) ;
- pâturage : lors du pâturage, les pissats et les bouses constituent un apport de nutriments non maîtrisable par l'agriculteur. Cette restitution au champ est accompagnée de phénomènes de volatilisation d'azote qui pourra éventuellement se redéposer ou contribuer à l'effet de serre (Hutchings et *al.*, 2007) ;
- récoltes de cultures, fauchage de prairies : les récoltes de cultures, et dans une moindre mesure les fauches de prairies, réduisent brusquement le couvert végétal du sol, ce qui le rend vulnérable à l'intensité des pluies et augmente les risques d'érosion (Houghton et Charman, 1986). Par ailleurs, ces opérations produisent des débris végétaux qui seront réintégrés au sol par le prochain travail du sol et par l'action des êtres vivants décomposeurs (micro-organismes, vers de terre, etc.) ;
- retournement de prairies et d'engrais verts<sup>21</sup> : cette opération combine les effets du travail du sol (fragmentation de la surface) et de l'enfouissement de débris végétaux (apport de nutriments et de

---

21 appelés aussi CIPAN (Culture Intermédiaire Piège À Nitrates), les engrais verts sont des cultures temporaires semées l'hiver et enfouies au printemps. Ce procédé permet d'immobiliser temporairement l'azote dans des tissus végétaux, limitant ainsi les risques de lessivage et de volatilisation pendant l'hiver et de le restituer dans les sols pour la culture suivante (Baggs et *al.*, 2000)

fibres végétales augmentant la stabilité structurale du sol) (Baggs *et al.*, 2000).

La lecture de l'itinéraire technique offre une vue synthétique de la ligne de conduite adoptée par l'agriculteur. En tant que séquence d'actions susceptibles de modifier les flux d'eau et d'éléments associés dans le paysage agricole, il constitue un support judicieux pour définir et préconiser des pratiques de gestion alternatives des cultures auprès des agriculteurs. L'établissement des rotations de cultures et de prairies joue donc un rôle clef dans l'enchaînement de ces itinéraires techniques et donc dans les programmes de protection de ressource en eau et des sols.

Castellazi *et al.*, (2007) ont proposé récemment un formalisme pour unifier la représentation des rotations de cultures, qui intègre approches séquentielles (série chronologique de cultures en boucle) et probabilistes (associant une probabilité à chaque transition entre deux cultures). Ces deux approches ont été utilisées pour optimiser la production des cultures, la préservation des sols (Dogliotti *et al.*, 2003; Rounsevell *et al.*, 2003; Bachinger et Zander, 2007) ou analyser les déterminants de la distribution spatiale des rotations (Mignolet *et al.*, 2007). En effet, le passage de l'échelle de la parcelle au bassin versant pose la question de la spatialisation des pratiques. Moussa *et al.* (2002) et Beaujouan *et al.* (2002) ont montré qu'elle influençait la réponse hydrologique du bassin versant, respectivement en terme de débit à l'exutoire et de flux de nitrates. Agarwal *et al.* (2002) ont proposé une typologie des modèles simulant l'évolution de l'occupation du sol, basée sur 3 axes de déterminants : les deux premiers sont l'espace et le temps, qui donnent le contexte de simulation des phénomènes biophysiques pris en compte, ainsi que la nature des processus humains de décision intervenant dans le changement d'occupation du sol. Les modèles dédiés à la simulation de succession de cultures fonctionnent généralement à l'échelle de la parcelle agricole, et les processus de décision de la culture à venir intègrent rarement plus que la culture précédente en terme d'historique et focalisent davantage sur les déterminants spatiaux (hydromorphie des sols, accessibilité des parcelles) et les objectifs de production de l'exploitation à laquelle appartient la parcelle.

### **Modélisation eau-sol-plante**

Cependant, la lecture des itinéraires techniques n'est pas suffisante pour appréhender le fonctionnement dynamique des sols et des cultures. En effet, la figure I.6 montre que l'absorption de différents nutriments varie dans le temps et que leurs dynamiques respectives sont fonction des cultures (Schvartz *et al.*, 2005). Ainsi, le développement des cultures modifie la teneur en nutriments du sol (ainsi que sa teneur en eau, par évapotranspiration). Par ailleurs, lorsque la fourniture du sol en nutriments est insuffisante (ou lorsque la teneur en eau libre est faible), des phénomènes de stress peuvent venir perturber le développement des cultures. Du fait de ces interactions, il est insuffisant d'appréhender la croissance des cultures seulement comme une réponse à l'état du système sol, mais il est nécessaire d'intégrer cultures et sols comme les éléments d'un même système dynamique (Wallach, 2006).

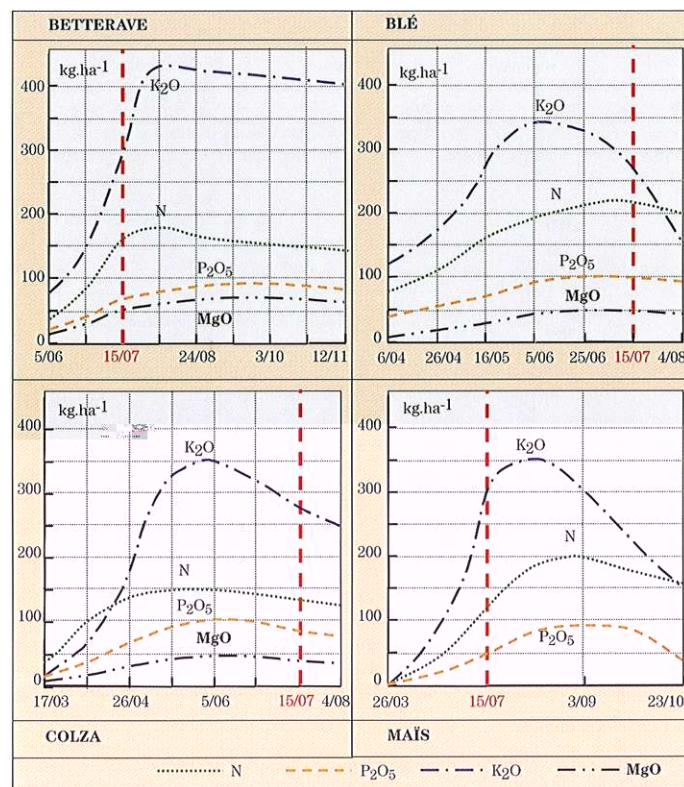


Fig. I.6: Courbes d'absorption au cours du cycle culturel de l'azote (N), du phosphore (P<sub>2</sub>O<sub>5</sub>), du potassium (K<sub>2</sub>O) et du magnésium (MgO) par les plantes entières de quelques cultures : betterave, blé, colza, maïs (Schvartz et al., 2005). Source des données : Les courbes d'absorption d'éléments minéraux. SCPA – Ministère de l'Agriculture et de la Pêche – DGER, 1990.

Il existe de nombreux modèles de culture qui intègrent les processus de transformation et d'assimilation de l'azote et du phosphore, souvent considérés comme les nutriments les plus limitants pour les cultures (Lewis et al., 2002 ; Gabrielle et al., 2002 ; Brisson et al., 2006). Ces modèles ont été conçus pour fonctionner à l'échelle de la parcelle et nécessitent des jeux de données conséquents pour initialiser et paramétrer la simulation : certains modèles font interagir plus de 10 formes moléculaires différentes pour les nutriments, chacune étant caractérisée par des coefficients de transformation vers une ou plusieurs autres formes. La spatialisation de ces modèles de cultures pose deux problèmes liés au changement d'échelle. Premièrement, l'« unknowability » des paysages ne permet plus de renseigner l'état initial de la simulation avec autant de précision (Beven, 2002). La spatialisation des modèles de culture STICS (Brisson et al., 1998) et EPIC (Williams et al., 1984), respectivement dans les modèles agro-hydrologiques TNT2 (Beaujouan et al., 2002) et SWAT (Grizetti et al., 2005), n'a pu se faire qu'au prix d'une simplification des modèles et d'hypothèses de stationnarité spatiale de certains paramètres. Leenhardt et al. (2006) suggèrent néanmoins des méthodes permettant de spatialiser l'information, notamment par l'utilisation de règles de pédo-transfert appliquées à diverses données obtenues par télédétection ; ou bien en effectuant des simulations préliminaires où les processus intégrés au modèle vont spatialiser eux-mêmes les résultats de simulation, qui pourront être alors utilisés comme état initial pour une autre simulation. Mais le deuxième problème dû à notre méconnaissance des paysages, est que la simulation des transferts de phosphore par ruissellement se heurte toujours à notre capacité réduite à simuler les mécanismes de déposition et de re-mobilisation des particules auxquelles le phosphore est adsorbé (Sonneweld et al., 2006, Deasy et al., 2008).

### ***Conclusion partielle sur les dynamiques hydrologique, pédologique et agricole des bassins versants***

Cette partie a permis de faire la synthèse sur la nature des processus d'évolution du paysage agricole. L'état hydrique et le couvert végétal du paysage jouent un rôle déterminant dans la genèse des flux connectant les précipitations au débit du cours d'eau. Cet état peut être fortement modifié par des interventions culturelles ponctuelles, mais aussi de façon plus continue par la croissance des cultures. C'est également en surface, dans les premiers horizons de sol, que les processus essentiels de transformation de la matière organique ont lieu. La spatialisation de leur intensité en fonction de l'hydromorphie des sols suggère à nouveau des interactions fortes avec le fonctionnement hydrologique du bassin versant. Les transferts de N et P sont associés au flux d'eau de ruissellement, de lessivage et de lixiviation qui suivent des dynamiques très différentes (plusieurs années pour les nappes, dans les bassins versants où les ruissellements atteignent le cours d'eau en quelques heures).

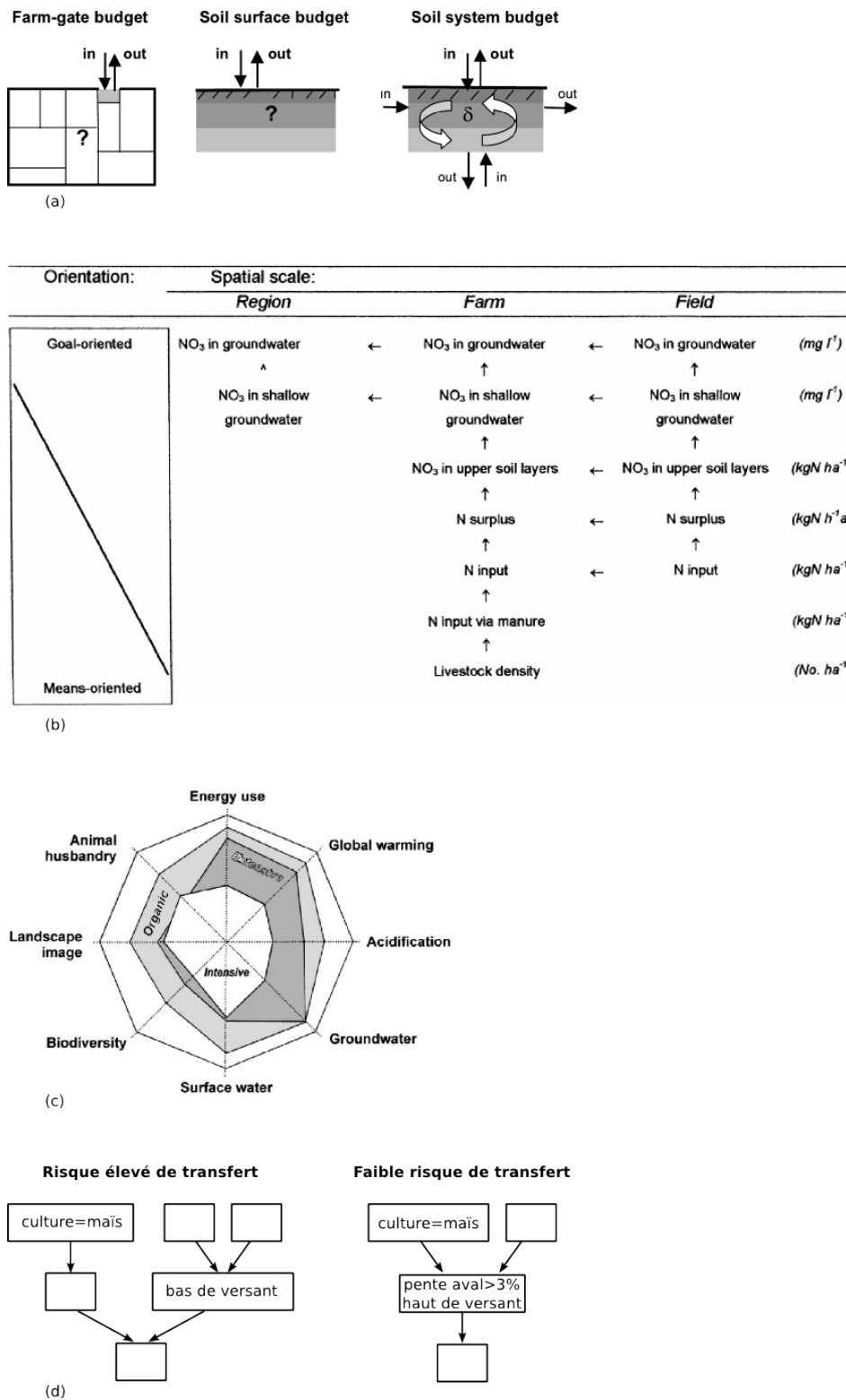


Fig. I.7 : Diversité d'indicateurs pour l'évaluation d'impacts environnementaux.

- (a) bilans de nutriments : différences conceptuelles de frontières et flux de nutriments pour les bilans exploitation, surface du sol, système sol (Oenema *et al.*, 2003) ;
- (b) indicateurs de fuite d'azote vers la nappe, échelle spatiale à laquelle ils interviennent. L'orientation d'un indicateur indique son aptitude à évaluer des pratiques (means) ou des objectifs à atteindre (goal) (Schröder *et al.*, 2003) ;
- (c) évaluation multi-critères de gestions de pâturage intensif, extensif, biologique par la méthode de l'Analyse de Cycle de Vie (ACV). Une valeur proche du centre du diagramme pour un critère représente un impact négatif (Haas *et al.*, 2001) ;
- (d) arbres d'exutoires de parcelles et caractéristiques associées à des risques de transfert de pesticides. Un carré vide représente une parcelle quelconque ; une flèche symbolise la relation « verse dans » (Trepas, 2008).

Compte tenu des dynamiques des différents processus d'évolution du paysage agricole (transfert, transformation), les impacts environnementaux des agrosystèmes peuvent être difficilement évalués sur une période de temps inférieure à la décennie, surtout lorsque l'on cherche à estimer l'effet de nouvelles pratiques agricoles. En effet, celles-ci modifient les contraintes auxquelles le système est soumis ; et le régime transitoire nécessaire à l'établissement d'un nouvel équilibre complique l'évaluation de l'impact final de ces pratiques. Enfin, les aléas climatiques introduisent une variabilité non maîtrisée dans les contraintes imposées au système qui le perturbent avec des inerties différentes : la génération des ruissellements de surface suivra la dynamique des événements pluvieux, le développement des cultures intègre les variations saisonnières, les fluctuations de nappe suivront le cycle hydrologique annuel.

## **Évaluation environnementale des agrosystèmes**

Par « évaluation environnementale », nous retiendrons ici la définition utilisée par l'Agence canadienne d'évaluation environnementale (2004), à savoir : il s'agit d'un processus ayant pour objet de prévoir les effets environnementaux d'initiatives proposées avant de les mettre en œuvre. Une évaluation environnementale n'est donc pas envisagée ici comme un contrôle (respect des cahiers des charges, justification de subvention), ni une étude d'impact *a posteriori* (teneurs en métaux lourds d'un terrain contaminé).

### **Indicateurs et méthodologies d'évaluation**

L'évaluation environnementale d'un agrosystème nécessite l'utilisation d'indicateurs qui, en résumant l'état ou le fonctionnement du système, va permettre d'en évaluer les impacts ou performances. Le choix des indicateurs employés par une méthode d'évaluation dépend notamment des limites de l'agrosystème étudié et des objectifs environnementaux auxquels elle s'intéresse, qui peuvent être classés dans trois catégories : ceux liés aux intrants, à l'état de l'agrosystème et aux facteurs d'émission (van der Werf et Petit, 2002 ; Payraudeau et van der Werf, 2005). La figure I.7 illustre les portées et les usages complémentaires de différents types d'indicateurs relatifs aux agrosystèmes.

Les trois types de bilan des nutriments proposés par Oenema et *al.* (2003) sont des indicateurs mono-variables permettant d'évaluer l'intensité et l'efficacité de l'utilisation d'un nutriment (Fig. I.7.a) :

- le bilan « exploitation » offre une vision intégrée de la pression exercée sur l'environnement en prenant compte de tous les flux entrant et sortant par les portes de l'exploitation. Cette approche est celle privilégiée dans la mise en œuvre de politiques incitatives, notamment parce que les données d'épandage, d'achat de fertilisants, de ventes des production sont relativement faciles à obtenir et avec une incertitude limitée (Öborn et *al.*, 2003) ;
- le bilan « surface du sol » prend en compte les flux de nutriments entrant par la surface du sol et qui la quittent par les récoltes, fauchage et pâture ;
- le bilan « système sol » prend en compte les mêmes entrées et sorties, plus les phénomènes de



recyclage et de transformation (minéralisation, assimilation, volatilisation) au sein du système, ainsi que les pertes en profondeurs. Cette approche, qui est davantage privilégiée en recherche dans les études sur le devenir des nutriments dans le sol, nécessite des phases lourdes d'acquisition de données, notamment pour suivre les variations saisonnières tout en couvrant un ou plusieurs cycles de culture (Öborn *et al.*, 2003).

Schröder *et al.* (2003) remarquent que les indicateurs liés aux pertes de nutriments (en l'occurrence, l'azote) dans un bassin versant agricole peuvent être hiérarchisés selon l'échelle spatiale à laquelle ils sont mesurés ou estimés, et en fonction de leur usage (Fig. I.7.b). Ainsi, à l'échelle du bassin versant, la concentration en nitrate d'un cours d'eau pourra être comparée au seuil maximal de 50 mg NO<sub>3</sub>.l<sup>-1</sup> fixé par la directive nitrate (European Commission, 1991), qui fournit ainsi un objectif de qualité à atteindre. Parallèlement, la directive exige que les bilans azotés fertilisation-récolte soient équilibrés au niveau de chaque exploitation, ce qui servira de référence pour évaluer les pratiques agricoles. Ainsi, du fait des phénomènes de transfert et de transformation des nutriments, plus un indicateur intègre des processus à une échelle spatiale vaste, plus l'information qu'il fournit s'éloigne des pratiques agricoles et de leurs effets directs (Schröder *et al.*, 2003).

L'évaluation environnementale des exploitations agricoles est progressivement devenue une étape de l'évaluation de leur durabilité, ce qui a soulevé deux défis. Le premier a été celui de passer à une évaluation intégrée, combinant plusieurs objectifs environnementaux et donc plusieurs indicateurs (environnementaux, écologiques, énergétiques, économiques, etc.). Le deuxième défi a été celui de s'affranchir de la spécificité de chaque exploitation pour pouvoir comparer la durabilité de modes de production. Pour cela, de nombreuses méthodes à l'instar de l'Analyse Cycle de Vie (ACV) ont exprimés ces impacts par unité fonctionnelle, comme par exemple le kilogramme de lait ou de viande produit, ou l'hectare d'exploitation (van der Werf et Petit, 2002). En combinant graphiquement ou mathématiquement ces indicateurs standardisés, il devient alors possible de faire une comparaison multi-critères des modes de production intensif, extensif et biologique pour un produit agricole (Fig. I.7.c). Le choix de l'unité fonctionnelle n'est cependant pas anodin : Halberg *et al.* (2005) montrent que le résultat de la comparaison de modes de production intensif et extensif s'inverse selon qu'on exprime l'impact par unité de produit fini ou de surface. À surface égale, le mode intensif impacte plus ; mais à production égale, le mode extensif impacte un territoire plus grand.

Cependant, si des méthodes comme l'ACV ont l'avantage de produire une image globale des impacts causés par un agrosystème, les démarches qu'elles utilisent pour caractériser le système sont généralement dépourvues de différenciation spatiale et temporelle – simplifications souvent jugées excessives pour des impacts concernant les pertes de nitrates et l'eutrophisation (Potting et Blok, 1994 ; Finnveden et Potting, 1999 ; Heijungs *et al.*, 2003). Ainsi, en couplant un modèle agro-hydrologique avec une ACV, Basset-Mens *et al.* (2006) ont remis en cause les hypothèses faites généralement dans les ACV, que le taux de transmission vers l'exutoire de l'azote en surplus sur les parcelles est constant et égal à 1. Également par l'analyse de résultats de simulation, Trepos (2008) a produit un autre type d'indicateurs, topologiques et sémantiques,

permettant d'associer des risques de transfert de polluants agricoles au cours d'eau à des configurations spatiales de parcelles et de pratiques agricoles (Fig. I.7.d).

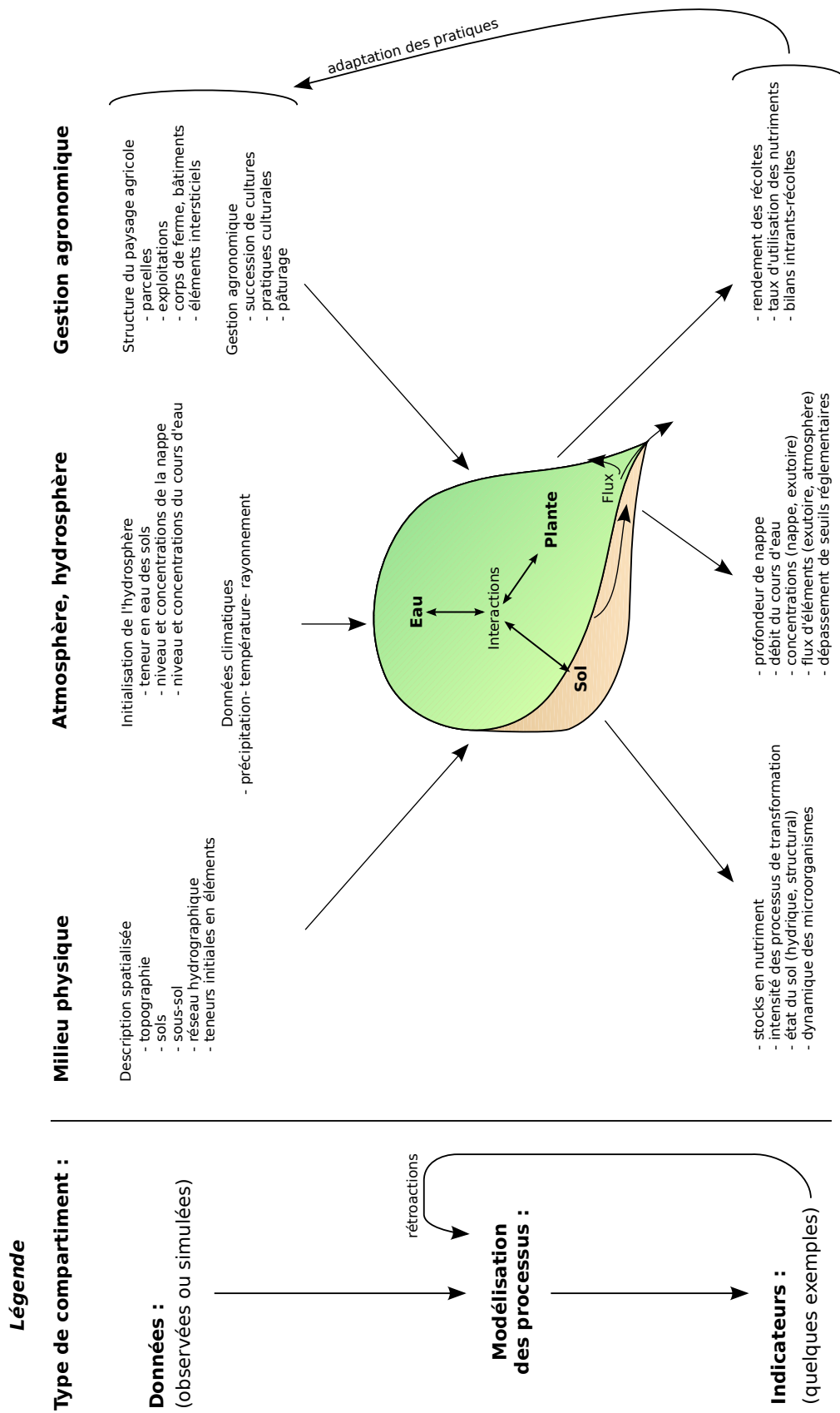


Fig. I.8 : Démarche de modélisation intégrée du paysage dans le cadre de l'évaluation des impacts environnementaux des agrosystèmes.

## **Modélisation et évaluation intégrée**

Les méthodes d'évaluation environnementale des agrosystèmes se sont complexifiées progressivement, tout d'abord en combinant des objectifs environnementaux différents (évaluation intégrée), puis comme l'ont montré les deux exemples précédents, par le couplage avec des modèles de simulation afin de tenir compte de la variabilité spatiale et temporelle des paysages au sein même du processus d'évaluation (évaluation et modélisation intégrées). En particulier, trois besoins justifient l'utilisation de la modélisation dans l'évaluation des impacts environnementaux des agrosystèmes :

- intégrer les différents processus d'évolution et leurs interactions, tester des hypothèses de fonctionnement pour l'analyse de sensibilité ou l'élaboration de nouvelles méthodologies ;
- surmonter les contraintes techniques et matérielles en simulant ce qui ne peut être mis en œuvre expérimentalement (réseau de mesures sur un territoire trop vaste, pendant une durée trop longue, climat non contrôlable, vision peu précise des activités humaines) ;
- construire et évaluer des scénarios alternatifs, notamment ceux impliquant une modification des pratiques agricoles, qu'on ne saurait imposer sans en connaître leur pertinence et leur viabilité pour les exploitations.

La figure I.8 propose une vue synthétique des flux d'information dans le cadre de l'évaluation des impacts environnementaux d'un paysage agricole. Les jeux de données relatifs au milieu physique, au climat et au territoire agricole peuvent être constitués soit par des relevés de terrain, soit par des approches de simulation qui devront alors faire preuve d'une certaine cohérence : respecter la sensibilité ou la tolérance des cultures à l'hydromorphie du sol, l'assolement du parcellaire avec les objectifs de production des exploitations, etc. Par contre, du fait des interactions relevées précédemment, la simulation des interactions eau-sol-plante doit être faite simultanément, notamment lorsque ces interactions ont des effets cumulatifs (effet du climat sur le développement végétal). La production d'indicateurs sur les résultats de simulation permet d'évaluer le fonctionnement du système et les impacts environnementaux. Ces indicateurs peuvent aussi être réinjectés dans le processus de simulation, notamment pour modifier les pratiques agricoles comme le font les agriculteurs quand ils adaptent l'assolement à venir de leurs parcelles en fonction des récoltes réalisées et des besoins de leurs élevages.

Les couplages de modèles envisagés dans le cadre de l'évaluation des impacts environnementaux des agrosystèmes soulèvent toutefois quelques défis (Van Asselt et Rotmans, 2002 ; Van Ittersum et *al.*, 2008) :

- les différents modèles ne sont pas forcément adaptés aux mêmes échelles spatiales ni au même pas de temps ; leur couplage conceptuel nécessite souvent la résolution de problèmes de transfert d'échelle ;
- ces problèmes d'échelle se traduisent généralement sur le plan technique par des complications sur la gestion des flux de données entre les modèles censés interagir à un pas de temps fin ou à une

résolution spatiale précise ;

- les différents modèles envisagés sont biaisés par le contexte dans lequel ils ont été développés, notamment lorsqu'ils sont spatialisés ou lorsqu'ils font intervenir des processus socio-économiques ;
- il est difficile d'estimer comment les incertitudes impliquées par les hypothèses simplificatrices de chaque modèle se combinent ni comment elles influencent le résultat final.

### **Plateforme de modélisation des paysages agricoles**

Le caractère unique d'un paysage et l'inconnaissabilité de sa structure (Beven, 2002) compliquent la comparaison de sites et le test d'hypothèses explicatives (Ruiz et *al.*, 2006). Si une hypothèse n'a pas pu être vérifiée expérimentalement, on pourra toujours trouver une explication possible à cet échec : effet du réseau de haies, de drains, fracturation du sous-sol, cheminements hydrauliques souterrains perturbés, etc. Le concept de paysage virtuel, qui est une représentation modélisée et donc simplifiée d'un paysage, a été développé comme solution à ce problème épistémologique dans de nombreuses disciplines telles que : en hydrologie (Beaujouan et *al.*, 2001 ; Beaujouan et *al.*, 2002), en pédologie (Walter et *al.*, 2003), en écologie (Ruiz et *al.*, 2006) et en gestion forestière (Gustafson et Crow, 1998). L'utilisation de paysages virtuels simplifiés permet de tester des hypothèses en fixant de façon délibérée des caractéristiques choisies du paysage, dans la mesure où le caractère simplificateur de la démarche ne compromet pas le test de l'hypothèse, ni l'interprétation des résultats.

Ainsi, le couplage de paysages virtuels avec les méthodes d'évaluation environnementale évoquées précédemment permet de comparer les impacts environnementaux de scénarios alternatifs. Afin de surmonter les défis scientifiques et les difficultés techniques liées à la modélisation et l'évaluation intégrées, des équipes de chercheurs multidisciplinaires ont récemment pris l'initiative de développer des plateformes de modélisation des agrosystèmes qui permettront de mutualiser les approches de création de paysage, de modélisation des processus, les outils de production et d'analyses d'indicateurs. En particulier trois projets parallèles tentent aujourd'hui de proposer une plateforme de modélisation des agrosystèmes à l'échelle du bassin versant, en suivant des approches similaires. Les plateformes SEVE<sup>22</sup> (Borrell et *al.*, 2005) et RECORD<sup>23</sup> (Chabrier et *al.*, 2007) sont des projets français visant respectivement à : (i) simuler les processus couplés impliqués dans le cycle de l'eau, du carbone, de l'azote et de substances polluantes aux échelles de la parcelle, du paysage et de la région, (ii) permettre l'analyse, l'évaluation et la conception de systèmes de culture innovants capables d'assurer des fonctions agronomiques et environnementales spécifiées. La troisième plateforme est un projet européen, SEAMLESS<sup>24</sup> (Van Ittersum et *al.*, 2008) dont l'objectif est de permettre l'évaluation de réglementations agricoles et environnementales sur des territoires allant de l'exploitation à l'Union Européenne. Ces trois plateformes sont toutes actuellement en phase de conception (RECORD, SEVE) ou de prototype (SEAMLESS) mais leurs cahiers des charges montrent des convergences

---

22 SEVE : Sol Eau Végétation Énergie

23 RECORD : RÉnovation et COORDination de la modélisation de cultures pour la gestion des agrosystèmes

24 SEAMLESS : System for Environmental and Agricultural Modelling; Linking European Science and Society

conceptuelles et techniques :

- un effort important de décomposition du paysage en objets élémentaires connectés entre eux notamment par des relations topologiques ;
- l'utilisation de modules de transfert pour simuler les flux entre les objets ainsi que leurs changements d'état. Il s'agit en fait de modèles déjà publiés que l'on souhaite modifier le moins possible pour alléger le travail de développement de la plateforme ;
- la mise à disposition pour l'utilisateur d'outils facilitant l'intégration de données réelles pour la représentation du paysage au sein de la plateforme, ou la création de paysages virtuels et de scénarios d'évolution alternatifs ;
- le développement d'un méta-programme, appelé « superviseur », qui coordonne le couplage dynamique des différents modules au cours de la simulation ainsi que les flux de données. Les plateformes souhaitant contraindre le moins possible la diversité des configurations que les utilisateurs voudront tester, la conception du superviseur semble être le point le plus délicat des trois plateformes ;
- la mise à disposition pour l'utilisateur d'outils produisant des indicateurs synthétiques pour faciliter l'analyse des simulations.

### ***Apport de la POO à la modélisation intégrée du paysage***

Comme le souligne Beven (2002), la plupart des scientifiques modélisateurs sont davantage des utilisateurs de modèles que des développeurs. Il semble cependant que les disciplines se rapprochent à leur avantage car de nombreux auteurs soulignent les intérêts de faire un lien direct entre les objets conceptuels du modèle et les objets informatiques du programme associé (Chabrier et *al.*, 2007). En effet, la Programmation Orientée-Objet (POO) est un paradigme de programmation où :

- les structures de données et les fonctions qui les manipulent sont regroupées en **objets**. Un objet est caractérisé par des *attributs* (valeurs numériques, chaînes de caractères, ou un autre objet) et des *méthodes* qui les manipulent (initialisation, incrémentation, calcul) ou qui permettent de connaître l'état des attributs (affichage de valeurs, création de fichier de sortie) ;
- il est possible de concevoir un nouvel objet par **héritage** en se basant sur un autre. Le nouvel objet hérite alors automatiquement des attributs et des méthodes de l'objet d'origine, auxquels on peut en rajouter d'autres (une « zone de terrain » peut être déclinée en « parcelle agricole » ou en « plan d'eau », tous les deux caractérisés par un même attribut de surface) ;
- par **polymorphisme**, il est possible de créer des objets différents qui auront des méthodes du même nom, ce qui facilite le traitement d'un même événement par des différents objets (le calcul du volume d'eau ruisselé lors d'un événement pluvieux selon la nature du terrain sur lequel on se trouve).

Outre les analogies sémantiques entre objet du paysage et objet informatique, la POO induit dans la pratique une grande modularité du code (chaque objet étant codé indépendamment des autres) qui facilite le développement et la maintenance du programme informatique final.

Les trois projets de plateforme SEVE, RECORD et SEAMLESS ont adopté une approche orientée-objet dans leur développement, non seulement pour les raisons techniques et sémantiques évoquées ci-dessus, mais aussi car ce mode de représentation fournit un cadre homogène pour articuler les différents modèles qui à disposition des utilisateurs de la plateforme.

## **Conclusion, problématique et démarche adoptée**

La synthèse des connaissances sur la modélisation intégrée du fonctionnement hydrologique et pédologique des paysages et des pratiques agricoles nous a permis de proposer un schéma de synthèse pour l'évaluation des impacts environnementaux des agrosystèmes (Fig. I.8). Il met en avant trois axes pour construire des paysages virtuels : le milieu physique qui contrôlera en partie les flux d'eau et d'éléments ; l'agencement des parcelles agricoles et les successions de cultures distribuent les pratiques agricoles dans l'espace et le temps ; enfin le climat est à la fois le moteur des flux hydrologiques et du développement des cultures. Les plateformes de modélisation et d'évaluation intégrée des agrosystèmes actuellement en développement ne sont pas encore opérationnelles, mais les formalismes qu'elles hébergeront sont pour la plupart accessibles sous la forme de modèles individuels, notamment en ce qui concerne la simulation spatialisée des interactions eau-sol-plantes ainsi que les flux d'eau et d'éléments associés.

En revanche, certains points ne sont pas encore clairement établis :

- comment simuler des paysages virtuels en assurant la cohérence entre milieu physique et les pratiques agricoles ?
- comment s'assurer que les paysages virtuels simulés rentrent dans les conditions de validité du modèle agro-hydrologique choisi ?
- quel est le délai de stabilisation du système pour les différents paysages virtuels construits ?
- quelle méthodologie multi-critères et multi-ressources mettre en œuvre pour distinguer dans les résultats de simulation les effets du climat, du milieu physique et des pratiques agricoles ?
- quels sont les acquis méthodologiques et les résultats transposables à un paysage réel ?

La problématique de ce travail d'évaluation d'impacts environnementaux réside donc dans la spatialisation des propriétés physiques et agricoles du paysage ; dans notre capacité à simuler l'effet de leur spatialisation sur la qualité de l'eau et des sols ; et enfin dans l'intégration des résultats de simulation pour pouvoir établir un diagnostic.

La démarche adoptée pour cette étude consiste à élaborer un modèle de représentation du paysage agricole qui permette dans un premier temps d'analyser, à partir de données collectées sur un petit bassin versant agricole breton, les déterminants spatiaux et temporels de l'allocation des cultures sur le parcellaire. Les différents paysages virtuels seront construits selon un plan factoriel combinant des variantes du milieu physique, du climat avec des pratiques agricoles alternatives sur le plan de l'occupation du sol et des pratiques de fertilisation N et P. Le modèle de représentation du paysage sera à nouveau utilisé afin de proposer une méthodologie originale de synthèse des résultats de simulation.

## **Chapitre II**

La plateforme *Qualscape*  
comme support de la modélisation des paysages agricoles





## **Chapitre II : La plateforme *Qualscape* comme support de la modélisation des paysages agricoles**

La plateforme *Qualscape* (*Landscape modeling to assess the impacts of agrosystem on natural resource quality*) vise à proposer une base conceptuelle et une démarche pour la simulation de paysages agricoles et l'évaluation environnementale des agrosystèmes. Les développements réalisés ont concerné la représentation du paysage agricole, la création de paysages virtuels et la simulation de leur évolution sous l'action de contraintes agronomiques (changement de l'occupation du sol, conduite de cultures) et climatiques. Nous présentons dans cette partie les développements méthodologiques et conceptuels mis en œuvre dans *Qualscape* et nous ferons référence aux chapitres III et IV pour détailler les applications réalisées et les résultats obtenus. Le site d'application choisi pour cette étude est le bassin versant de Naizin situé dans le Morbihan (France).

### **Méthodologie générale – de la création de paysages virtuels agricoles à leur évaluation environnementale**

L'état des connaissances réalisé dans le chapitre précédent nous a permis d'identifier trois facteurs que nous avons utilisés pour construire un jeu de paysages virtuels agricoles : le climat (à l'origine des flux d'eau), le milieu physique (support des cultures et des flux) et le système de culture (déterminant l'occupation du sol et les apports d'intrants). Trois modalités de milieu physique ont été créées en combinant des épaisseurs et des caractéristiques de sols différentes. Enfin, trois séries climatiques différentes ont été créées à partir d'une chronique observée, pour obtenir une série climatique plus sèche et une autre plus humide. Deux modalités de systèmes de cultures, différant sur les pratiques de fertilisation et la gestion de l'interculture hivernale, ont été définies pour caractériser un mode de production intensif non raisonné et un mode plus raisonné où des intercultures d'hiver sont plantées pour protéger les sols nus et piéger les nitrates dans la biomasse, et où les apports d'azote et de phosphore ont été réduits. Dix huit paysages virtuels résultants de la combinaison de tous ces facteurs ont été construits (Fig. II.9).

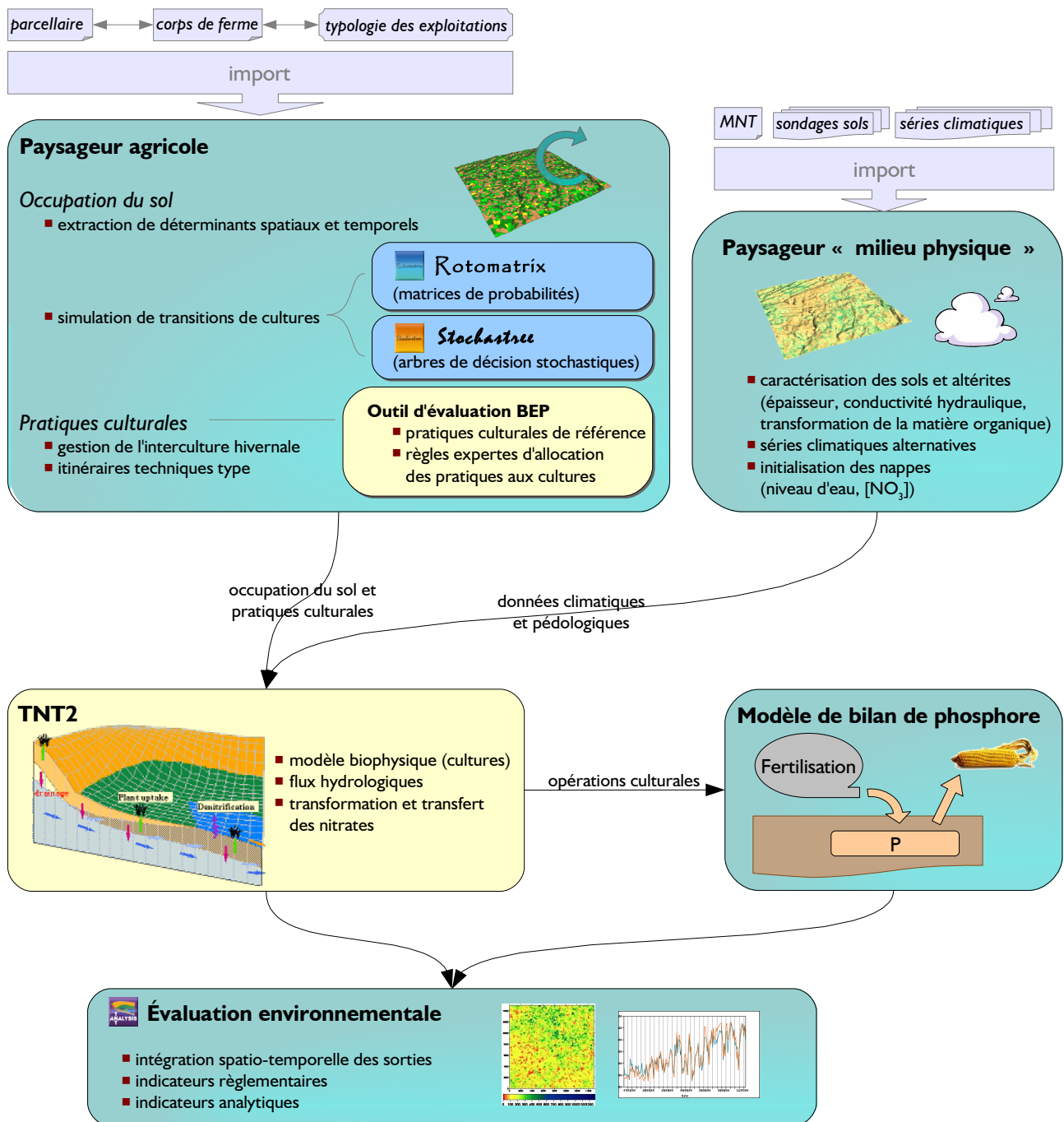


Fig. II.10 : Démarche générale adoptée dans *Qualscape* pour la construction de paysages virtuels (en haut de la figure) et l'évaluation environnementale des agrosystèmes (en bas de la figure). Les flèches représentent les flux de données entre les modules (en jaune, les modules existants intégrés dans notre démarche).

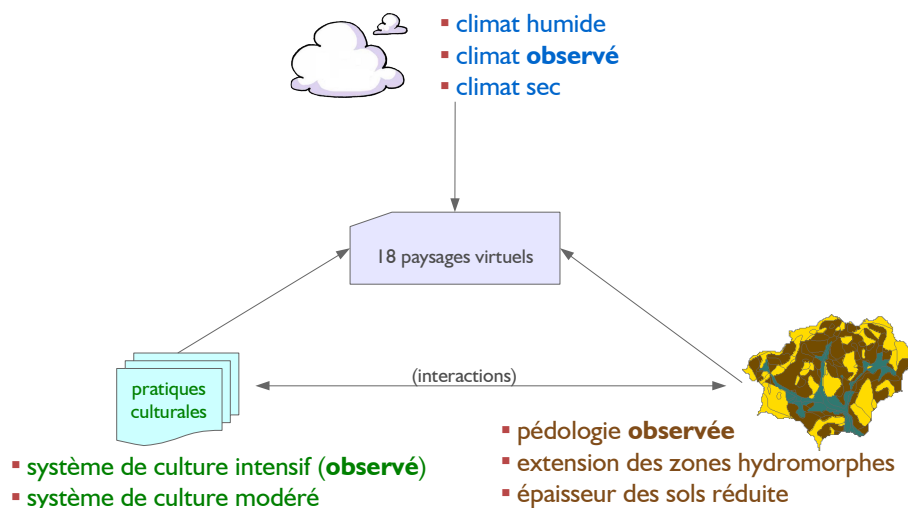


Fig. II.9 : Les combinaisons de facteurs climatiques, de systèmes de culture et du milieu physique qui ont permis la création des 18 paysages agricoles virtuels.

L'occupation du sol conditionnant la spatialisation des pratiques agricoles et leur succession dans le temps, une première étape nécessaire a été d'analyser, à partir d'un jeu de données observées, les déterminants spatiaux et temporels des transitions de culture, puis à les intégrer en tant que contraintes agronomiques dans un modèle de transition de cultures basés sur des arbres de décision stochastiques (*Stochastree*). Un modèle plus simple, basé sur des matrices de probabilités de transition de cultures a également été développé (*Rotomatrix*).

La démarche d'analyse des déterminants de l'occupation du sol, la construction du modèle ainsi que l'évaluation de ces performances est détaillée dans le **chapitre III** de ce rapport.

Cette étape a permis de conclure que *Stochastree* permettait de simuler les transitions de cultures en adéquation avec les déterminants spatiaux évoqués précédemment. C'est donc le modèle que nous avons utilisé pour simuler l'évolution de l'occupation du sol en adéquation avec les différentes modalités de milieu physique.

La construction des 18 paysages virtuels ainsi que les résultats de leur évaluation environnementale sont détaillées dans le **chapitre IV** de ce rapport.

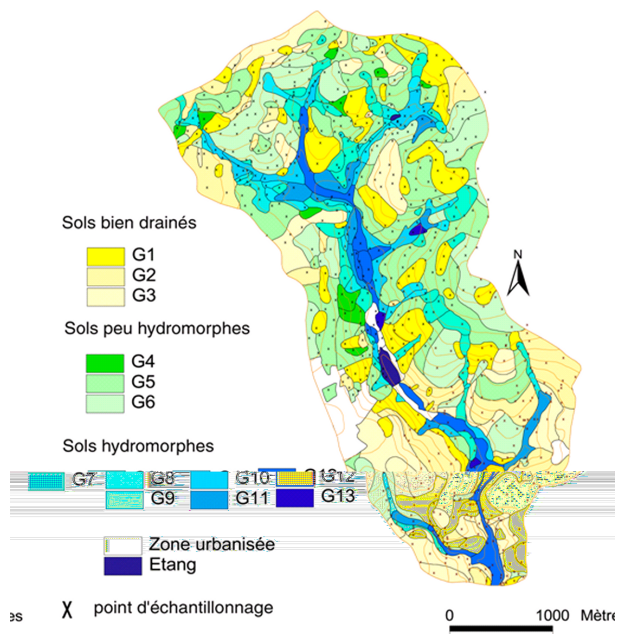
La figure II.10 schématise la démarche générale adoptée dans le projet. Le *paysageur* « milieu physique » regroupe les étapes ayant permis de construire les différents milieux physiques et chroniques climatiques. Le *paysageur agricole* regroupe les étapes ayant permis d'étudier et de simuler l'évolution de l'occupation du sol, et d'associer à chaque culture<sup>25</sup> des pratiques culturelles (semis, fertilisations, récolte<sup>26</sup>). La combinaison des données relatives aux systèmes de culture, au climat et au milieu physique constitue le jeu de données d'entrée au modèle biophysique et hydrologique *TNT2* (Beaujouan et al., 2002). Les sorties de *TNT2*

25 le terme de culture est entendu dans une acception large, intégrant aussi les prairies

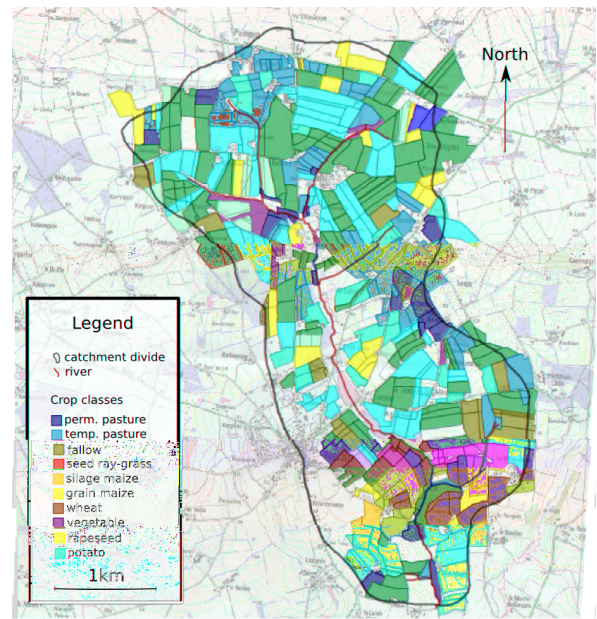
26 le terme de récolte est entendu dans une acception large, intégrant aussi les exports par pâturage et fauchage de prairies

relatives aux pratiques culturales (notamment les fertilisations et les récoltes) servent d'entrées à un *modèle de bilan entrées-sorties de phosphore* dans les sols. Les sorties de *TNT2* et de ce modèle phosphore sont alors traitées afin de dresser l'*évaluation environnementale du paysage agricole* virtuel.

Dans la suite de ce chapitre, nous présentons les formalismes que nous avons développés concernant la représentation du paysage agricole, en particulier du parcellaire. Le premier, relatif à l'occupation du sol, a pour objectifs (i) d'intégrer les déterminants spatiaux et temporels identifiés par l'analyse des successions culturales réalisées dans le chapitre III, (ii) de servir de base à une modélisation orientée-objet des transitions de cultures. Le deuxième, relatif à l'analyse des pratiques culturales, a pour objectifs (i) d'offrir une représentation-objet des sorties parcellaires de *TNT2*, (ii) de faciliter le calcul de bilans d'azote, (iii) de servir de support pour le calcul de bilans de phosphore dans le sol des parcelles du paysage. Nous avons choisi d'implémenter ces formalismes dans le langage informatique Java, qui offre un accès relativement simple à la programmation orientée-objet et qui dispose de bibliothèques nous offrant une interopérabilité avec les différents format de données que nous avons manipulés (MySQL, CSV, XML, Shapefile ESRI, etc.).



(a)



(b)

Fig. II.11 : Cartes des sols et du parcellaire du bassin versant de Naizin (Morbihan) :

- (a) carte des sols identifiant 13 grands groupes de sols et positionnant les sondages (Walter et Curmi, 1998) ;
- (b) carte du parcellaire représentant l'occupation du sol en 2000.



# Formalisme du paysage « occupation du sol » dans Qualscape

## ***Le bassin versant de Naizin, support de la démarche***

Les travaux méthodologiques développés dans Qualscape ont été fondés sur un bassin versant réel qui a servi de support aux approches de modélisation : il s'agit du bassin versant de Naizin (Morbihan, France) qui fait l'objet de travaux de recherche depuis 1971 sur l'évolution des paysages et de la qualité de l'eau, qui ont été synthétisés dans une monographie (Cheverry, 1998). Ce bassin versant comprend actuellement un des sites de l'Observatoire de Recherche en Environnement (ORE) dédié aux agro-hydrosystèmes (Agrhys) : l'ORE Agrhys a été labellisé par le Ministère de la Recherche en 2002 et a pour objectif d'étudier les temps de réponse des flux hydrogéochimiques à l'évolution des systèmes de production et des pratiques agricoles.

L'ensemble du bassin versant de Naizin tel que considéré dans le chapitre III couvre 1193 ha dont 85 % sont situés sur la commune de Naizin et 15% sur les communes de Réguiny et Kerfourn. Son point le plus haut culmine à 136 m et son exutoire se situe à 60 m.

Les sources d'information existantes sur le bassin versant sont nombreuses tant sur le suivi hydrologique, les caractéristiques du milieu (sols, climat), ainsi que sur les pratiques agricoles. Trois jeux de données ont été Notamment utilisés dans le chapitre III :

- une cartographie détaillée des sols à l'échelle du 1/10 000 (Fig. II.11.a) fondée sur 960 sondages à la tarière décrits avec précision et sur 28 profils pédologiques (Walter et Curmi, 1998) ;
- un relevé au sol pour la quasi-totalité des parcelles (Fig. II.11.b) des cultures d'été ; cet inventaire a été effectué entre 1993 et 2002 par le Cemagref (Bordenave et *al.*, 2005) ;
- en 2003, une enquête auprès des agriculteurs pour appréhender de façon précise les pratiques agricoles et leurs distributions à l'échelle du bassin versant (Payraudeau et *al.*, 2006).

Le couplage avec le modèle hydrologique *TNT2* (chapitre IV) a été réalisé sur un sous-bassin au nord du bassin versant global : il s'agit du bassin versant de Kervidy d'une superficie de 482 ha, qui correspond strictement à l'extension du site de l'ORE Agrhys<sup>27</sup> et qui, à ce titre, bénéficie d'une caractérisation hydrochimique en continue des nappes et des cours d'eau.

Le bassin versant de Kervidy est le siège d'une agriculture très intensive. En 1996, les céréales, maïs, prairies et temporaires et autres cultures occupaient respectivement 24%, 23%, 22% et 31 % de la surface totale. La charge en animaux est particulièrement forte, puisque à cette même date, le nombre de porcs par hectare était de 24, à comparer aux densités moyennes de 5,5 et 0,6 porcs/ha à l'échelle de la Bretagne et de la France.

---

27 site web officiel : <http://www.inra.fr/oreagrhy>



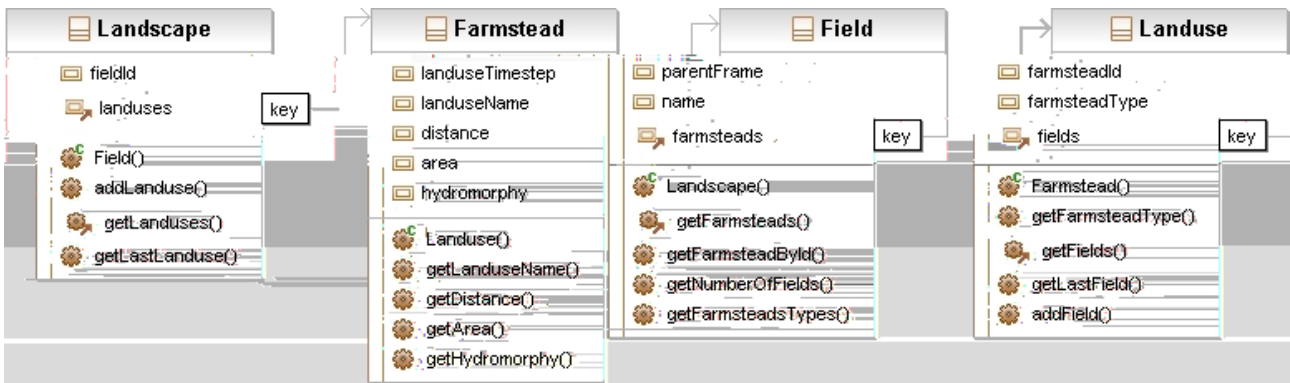


Fig. II.12 : Diagramme des classes utilisées pour représenter un paysage agricole dans *Qualscape*, dans le cadre de la simulation de transition de culture. Les en-têtes en gras sont les noms des classes. Dessous suivent les attributs des classes, puis les méthodes, figurées par des engrenages (un C vert symbolise la méthode « constructeur » de la classe). Par souci de clarté, nous n'avons représenté qu'une sélection d'attributs et de méthodes intervenant dans la description du paysage agricole. Les flèches précédées du mot « key » signifient ici qu'un attribut de la classe est une liste d'objets appartenant à la classe pointée par la flèche.

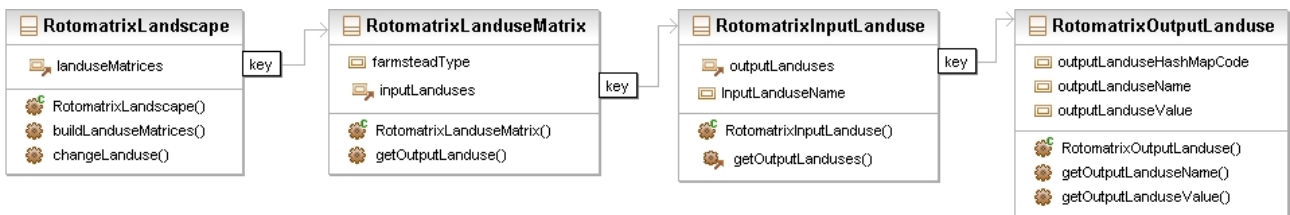


Fig. II.13 : Diagramme de classes utilisées par le modèle *Rotomatrix*. La classe *RotomatrixLandscape* dérive de la classe *Landscape*, en rajoutant une représentation des matrices de transition de cultures.

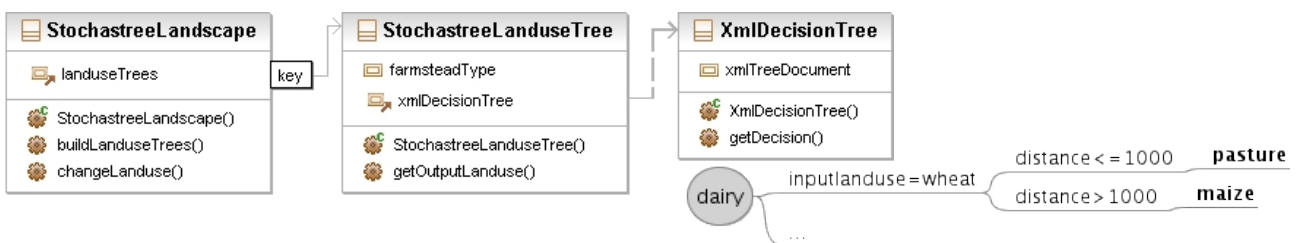


Fig. II.14 : Diagramme de classes utilisées par le modèle *Stochastree*. La classe *StochastreeLandscape* dérive de la classe *Landscape*, en rajoutant une représentation des arbres de décision. Le constructeur de la classe *XmlDecisionTree* requiert un fichier XML représentant la structure logique de l'arbre de décision. Un extrait d'un tel arbre est donnée sous forme graphique.

## **Une représentation hiérarchique du paysage agricole**

La figure II.12 représente le diagramme de classes utilisées pour la représentation d'un paysage agricole dans *Qualscape*. La classe englobante est *Landscape* (le paysage agricole en entier), dont l'attribut *farmsteads* est une liste d'objets de la classe *Farmstead*. De même, la classe *Farmstead* (symbolisant une exploitation agricole) contient une liste d'objets de la classe *Field* (une parcelle) dont l'ensemble constitue le parcellaire d'une exploitation. Enfin, la classe *Field* contient une liste d'objets de la classe *Landuse*, chacun décrivant l'occupation du sol à un moment donné. Ainsi, l'emboîtement des objets  $Field \in Farmstead \in Landscape$  propose une description spatiale du territoire agricole, et la succession de *Landuse* au sein de chaque *Field* propose une représentation temporelle des successions de culture.

L'étude des successions culturelles réalisées dans le chapitre III a montré l'importance du précédent cultural dans la détermination de la culture à suivre. Le modèle *Rotomatrix* a été conçu pour exploiter cette relation sous forme de matrices de probabilités de transition de cultures afin de simuler les successions de cultures. Quant au modèle *Stochastree*, il combine ce déterminant temporel avec des caractéristiques spatiales (hydromorphie de la parcelle, surface de la parcelle, distance de la parcelle au corps d'exploitation, type d'exploitation auquel est rattachée la parcelle) pour simuler les successions de cultures. Ainsi, la classe *Farmstead* possède un attribut *farmsteadType* pour identifier son type d'activité. Concernant les caractéristiques parcellaires, nous avons fait le choix de les intégrer dans la classe *Landuse* plutôt que *Field* afin de garder une trace d'éventuelles modifications de ces caractéristiques dans le temps. En effet, on peut envisager que la mise en place de drains modifie l'hydromorphie de la parcelle, qu'un échange de parcelle entre exploitation modifie la distance au corps d'exploitation ou qu'un agrandissement des contours en modifie la surface. Le code développé permet de modifier ces propriétés au sein d'une simulation mais nous n'avons pas exploité ces fonctionnalités dans ce projet.

Les deux parties suivantes présentent comment les modèles *Rotomatrix* et *Stochastree* ont été implémentés en tant que classes dérivées de la classe *Landscape*, bénéficiant ainsi d'une structure commune de représentation du paysage. Chacun d'eux dispose d'une interface graphique permettant d'initialiser un paysage par l'import d'un fichier DBF ou d'un shapefile<sup>28</sup>, de simuler l'évolution de l'occupation du sol et d'exporter le résultat sous forme de fichiers SIG. *Qualscape* dispose d'une boîte-à-outils d'entrées-sorties (*IO-toolbox*), conçue pour gérer tous les transferts de données requis par la démarche.

## **Spécifications de la représentation du paysage pour le modèle Rotomatrix**

Le modèle *Rotomatrix* simule les successions de cultures en se basant sur des probabilités de transitions de cultures structurées en matrice : les différentes lignes correspondant aux cultures en place et les colonnes les différentes cultures possibles à venir. Les matrices de probabilités utilisées dans le chapitre III sont présentées en annexe II. L'utilisation de fichiers texte au format CSV<sup>29</sup> a été utilisée comme format pour

28 DBF : database file. Il s'agit du format utilisé pour le stockage de la table d'attributs des fichiers Shapefile ESRI, utilisés dans de nombreux systèmes d'information géographique (SIG)

29 CSV : comma-separated value

représenter ces données. La classe *RotomatrixLandscape* (Fig. II.13) possède une méthode *buildLanduseMatrices()* qui, à partir d'un chemin d'accès donné en paramètre, va construire autant de matrices de transition qu'elle trouve de fichiers appropriés dans le répertoire donné : les fichiers pris en compte doivent porter l'extension « *.matrix* ». Le préfixe du nom de fichier est utilisé pour identifier le type d'exploitation auquel la matrice se rapporte, et sert à renseigner l'attribut *farmsteadType* de la classe *RotomatrixLanduseMatrix*. L'attribut *inputLanduses* est une liste d'objets de la classe *RotomatrixInputLanduse* et correspond aux cultures « en cours » des différentes lignes de la matrice (le nom de la culture en cours est stocké dans l'attribut *inputLanduseName*). La classe *RotomatrixInputLanduse* contient une liste d'objet de la classe *RotomatrixOutputLanduse*, caractérisée par deux attributs relatifs aux « cultures suivantes » décrites par les colonnes de la matrice : l'attribut *outputLanduseName* est le nom de la culture à suivre, *outputLanduseValue* porte la probabilité que cette culture succède à une culture en place donnée.

Dans le processus de simulation de transition de culture, l'appel de la méthode *changeLanduse()* de la classe *RotomatrixLandscape* va :

- parcourir chaque exploitation (classe *Farmstead*) du paysage et en récupérer le type d'activité avec la méthode *getFarmsteadType()* (Fig. II.12) afin de savoir quelle matrice utiliser ;
- pour chacune des parcelles (classe *Field*) d'une exploitation, récupérer l'occupation du sol en cours avec la méthode *getLastLanduse()* ;
- utiliser les probabilités de transition de la matrice correspondant au type d'exploitation pour déterminer de façon stochastique la culture à suivre.

À la fin d'un appel de *changeLanduse()*, toutes les parcelles ont une occupation du sol de plus dans leur liste de *Landuse*. Si une parcelle contient une occupation du sol qui n'est pas gérée par la matrice, la nouvelle occupation est une copie de la première. Cela permet de gérer automatiquement des occupations du sol considérées comme pérennes à l'échelle de la simulation, comme les routes, le bâti, les forêts et les plans d'eau.

### **Spécifications de la représentation du paysage pour le modèle Stochastree**

Le modèle *Stochastree* intègre les différents déterminants de l'occupation du sol identifiés dans le chapitre III sous la forme d'arbres de décision, ce qui a nécessité une tout autre approche pour la simulation des transitions de culture. Les informations manipulées par les arbres de décision sont des couples attribut-valeur caractérisant une situation, et qui vont être testés aux différents nœuds de l'arbre. Si le test est positif, on progresse dans l'arborescence jusqu'à atteindre une feuille terminale qui renvoie la décision correspondant à la situation décrite. Si le test est négatif, on passe au test voisin se trouvant au même niveau de progression dans l'arbre. Dans le cas de *Stochastree*, les couples attribut-valeur testés par les arbres de décision sont les caractéristiques d'une parcelle (culture en place=*i*, hydromorphie=*n*, distance au corps d'exploitation=*x*,

surface=y), et la décision attendue est la réponse à la question « quelle est la prochaine culture ? ». Un extrait simpliste d'un arbre de décision est donné dans la figure II.14 . La racine de l'arbre y est représenté à gauche et le reste de l'arbre se lit de gauche à droite. Ainsi, dans cet exemple :

- le premier attribut testé par l'arbre est *inputlanduse* (culture en place) ;
- s'il s'agit d'un blé (*inputlanduse=wheat*), l'attribut suivant testé est la distance de la parcelle au corps d'exploitation ;
- si cette distance est inférieure à 1000 m, alors l'arbre « propose » comme culture suivante une prairie (*pasture*) ; si la distance est supérieure à 1000 m, c'est un maïs qui est suggéré.

Plutôt que de chercher à développer une représentation-objet des arbres de décision, nous avons défini un format XML<sup>30</sup> pour les représenter, et avons développé la classe *XmlDecisionTree* qui gère le parcours de tests dans un arbre XML suivant ce format. L'annexe III présente le format XML que nous avons défini dans ce projet.

La classe *StochastreeLandscape* (Fig. II.14) contient une liste d'objet de la classe *StochastreeLanduseTree*, qui associe un arbre de décision (attribut *xmlDecisionTree*) à un type d'exploitation (attribut *farmsteadType*). La méthode *buildLanduseTrees()* permet d'automatiser la création de cette liste de *StochastreeLanduseTree* à partir des fichiers terminés par l'extension « *.tree.xml* » localisés dans un même répertoire. La méthode *changeLanduse()* de la classe *StochastreeLandscape* a un comportement similaire à celle de *RotomatrixLandscape* : elle parcourt les parcelles de chaque exploitation et utilise les caractéristiques de la parcelle pour déterminer la culture à venir. Le fonctionnement stochastique des arbres est détaillé dans le chapitre III et son implémentation dans la fin de l'annexe III.

## **Conclusion sur la modélisation de paysages agricoles dans Qualscape**

À partir d'un formalisme volontairement générique de la structure du paysage agricole, nous avons dérivé deux approches de simulation des transitions de culture. L'originalité de notre approche réside en particulier dans le couplage fin entre les objets du paysage et les processus d'évolution de l'occupation du sol. Ainsi, en fin de simulation, nous disposons d'une description spatio-temporelle du paysage, dont la structure peut être exploitée à des fins analytiques : nous pouvons ainsi étudier les successions culturelles au niveau de la parcelle, de l'exploitation, du type d'exploitation ou du paysage. Par ailleurs, le choix d'un formalisme simple pour la représentation du paysage en facilite le couplage avec des SIG, pour la création de l'état initial du paysage, comme pour l'analyse des sorties.

---

30 XML (eXtended Markup Language) est un langage à balise dont la syntaxe se prête particulièrement bien à la représentation de structures arborescentes

# Formalisme du paysage « pratiques culturelles » dans *Qualscape*

## Les sorties parcellaires du modèle *TNT2*

Le modèle *TNT2*, présenté plus en détail dans le chapitre IV, combine un modèle biophysique spatialisé simulant les interactions eau-sol-plante sous les contraintes climatiques et au rythme des opérations culturales réalisées sur chaque parcelle. *TNT2* génère des fichiers de sorties parcellaires récapitulant les différentes opérations culturales qui y ont eu lieu. L'exemple suivant est un extrait montrant les opérations culturales, du semis à la récolte, réalisées sur une parcelle :

```
Debut de l'ITk m_gLP2_p1 sur les parcelles no 104 au jour 607 (19/04/2002)
20/04/2002 (j=608) semis mais
21/04/2002 (j=609) ferti mais N mineral apporte(kgN/ha) = 147.10
22/04/2002 (j=610) ferti mais N mineral apporte(kgN/ha) = 25.00
14/10/2002 (j=785) recol mais N exporte(gN/m2) = 9.93 [suite à la ligne]
[...] Norg restitue(gN/m2) = 11.33 MS aerienne (t/ha) = 20.05 dont grains (t/ha) = 10.62
```

On y distingue notamment :

- la quantité d'azote apportée par hectare de parcelle ;
- la quantité d'azote et de matière sèche exportée par les récoltes.

Afin d'exploiter ces résultats de simulation à différentes échelles spatiales et temporelles, un autre formalisme de paysage agricole a été développé.

## Paysage « pratiques culturelles »

La figure II.15 présente la structure du paysage « pratiques culturelles », qui est construit sur une base similaire au paysage « occupation du sol ». La classe englobante *CulturalPracticeLandscape* dérive de la classe *Landscape* et y rajoute notamment l'attribut *parser*, qui permet de convertir les sorties parcellaires de *TNT2* en objets de la classe *CulturalPractice*. La classe *CulturalPracticeField* dérive de la classe *Field* et y rajoute notamment comme attribut une liste de *CulturalPractice* qui sera alimentée par le *parser*.

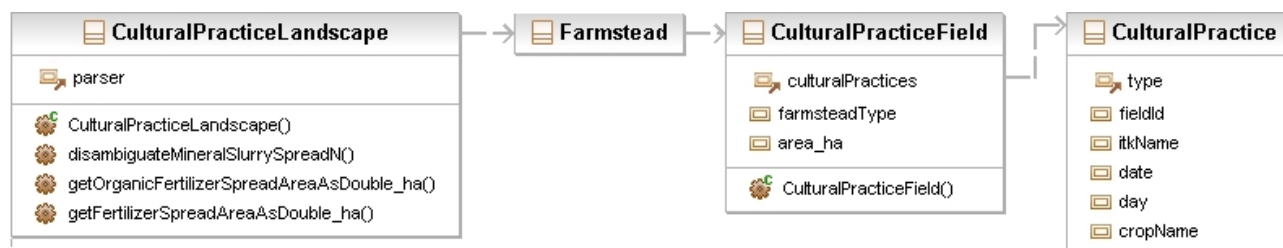


Fig. II.15 : Diagramme des classes utilisées pour représenter un paysage agricole dans *Qualscape*, dans le cadre de l'analyse de pratiques culturelles. Par souci de clarté, nous n'avons représenté qu'une sélection d'attributs et de méthodes permettant l'analyse des pratiques culturelles.

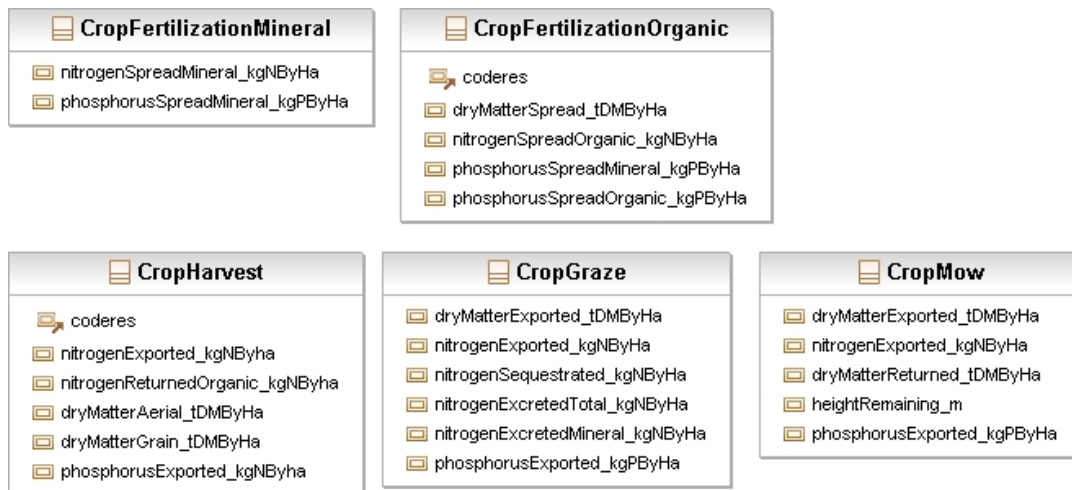


Fig. II.16 : Les différentes classes de pratiques culturales dérivant de la classe *CulturalPractice*. Les classes *CropFertilizationMineral* et *CropFertilizationOrganic* symbolisent respectivement les fertilisations par engrais de synthèse et par épandages d'effluent. *CropHarvest*, *CropGraze* et *CropMow* caractérisent les exports de matière organique (N, P, matière sèche) et les résidus dus aux récoltes, au pâturage et au fauchage.

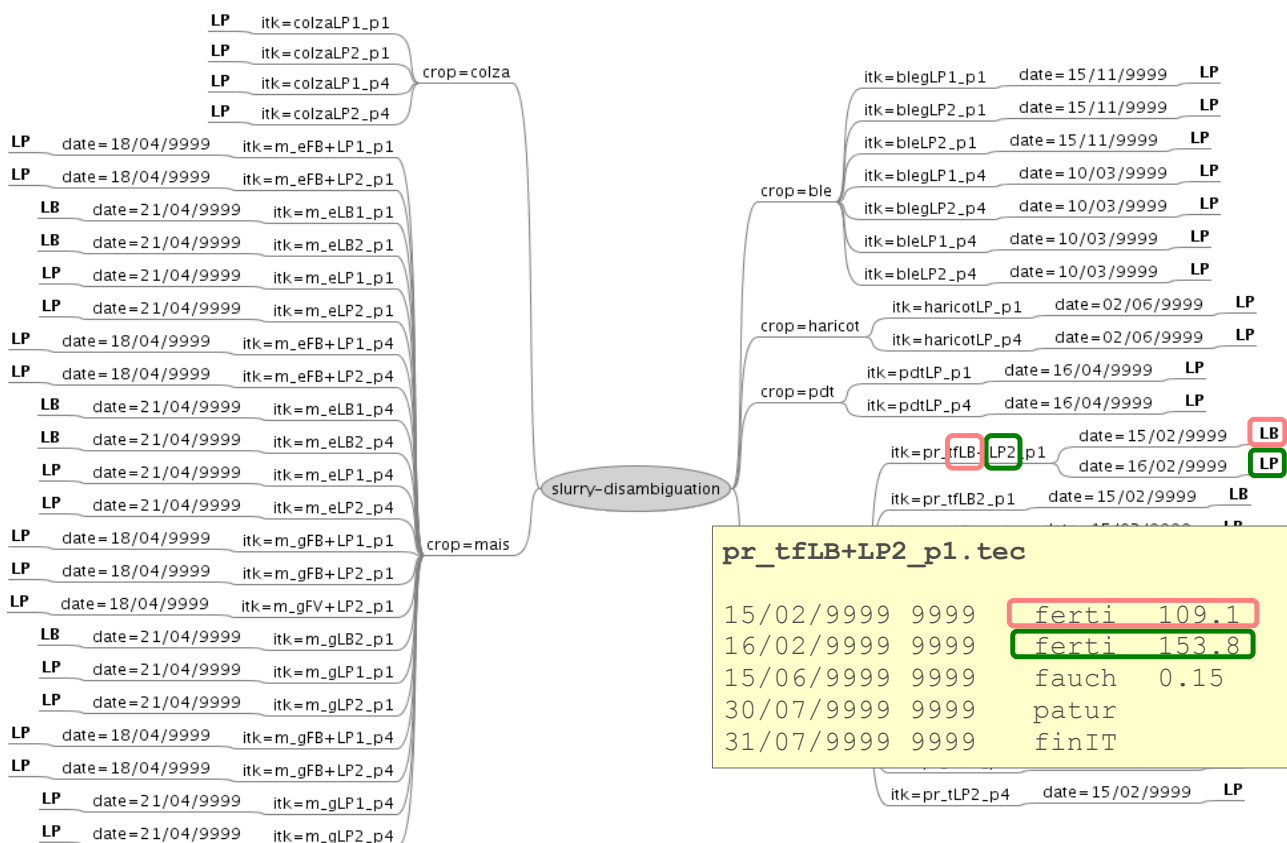


Fig. II.17 : Différencier les apports d'azote par lisier et par engrais minéral en testant les itinéraires techniques type avec un arbre de décision. L'itinéraire technique présenté (encadré jaune) correspond à la gestion d'une prairie comprenant deux fertilisations (les valeurs encadrées sont en kg N.ha<sup>-1</sup>). L'arbre de décision (en arrière-plan) permet de déterminer, via les noms des itinéraires techniques et les dates de fertilisation, la nature des effluents épandus (LP : lisiers de porcs, LB : lisier de bovin).

La classe *CulturalPractice* sert de structure générique pour stocker des informations relatives à la culture en place, au nom de l'itinéraire technique en cours et à la date à laquelle s'effectue la pratique. De cette classe dérivent des pratiques culturelles spécifiques aux fertilisations et aux exports-résidus par récolte, pâturage et fauchage (Fig. II.16). Une classe *CropSow* a également été créée pour garder trace du semis de la culture, mais nous n'en avons pas trouvé d'usage spécifique (*TNT2* ne simulant pas l'effet de ces pratiques sur l'état de surface du sol).

L'utilisation d'un arbre de décision combinant le nom des itinéraires techniques et les dates des opérations culturales nous a permis de différencier les apports de lisiers des fertilisations minérales (simulés comme des fertilisations minérales dans *TNT2* du fait de l'état fortement minéralisé de l'azote dans les lisiers) (Fig. II.17).

### **Modèle phosphore**

Il se base sur un paysage « pratiques culturales » et utilise des ratios P/N et P/MS (matière sèche) pour simuler les apports-exports de P via les fertilisations et les récoltes/pâturages/fauches. Son fonctionnement est détaillé dans le chapitre IV.

Paramètre TNT	t_sols0	t_sols1	t_sols2
	hydromorphe	intermédiaire	drainé
1 - transmissivité latérale ( $m^2/j$ )	1,8	6,0	6,0
2 - facteur de décroissance exponentiel ( $m$ )	0,001	0,08	0,08
7- macro-porosité du niveau intermédiaire ( $cm^3/cm^3$ )	0,15	0,2	0,2
8 - rapport des porosités (micro/macro) du niveau intermédiaire (-)	1,2	1,05	1,05
9 - épaisseur d'altérite ( $m$ )	12,	25,	25,
20 - teneur initiale en azote organique de l'humus ( $g/100 g sol$ )	0,27	0,25	0,20
21 - rapport C/N de l'humus (-)	15	12	10
22 - fraction du stock d'humus inactive (-)	0,45	0,48	0,60
23 - vitesse potentielle de dénitrification ( $g.m^{-2}.j^{-1}$ )	1,6	0,8	0,8

Table II.1 : Paramètres de *TNT2* dont les valeurs différencient les trois types de sols identifiés sur Naizin.

## Représentation du milieu physique

Le milieu physique constitue le support de croissance des cultures, mais aussi celui des flux d'eau et des éléments associés. Dans ce projet, ces processus sont simulés par le modèle *TNT2*, nous avons donc décrit le milieu physique du paysage en fonction du formalisme suivi par ce modèle. Les propriétés des sols ont été spatialisées horizontalement selon des grilles régulières constituées de pixels de 20 m de côté, selon des méthodes qui seront présentées dans cette partie.

### **Spatialisation de l'épaisseur des sols**

Concernant l'épaisseur des sols, nous disposons d'un jeu de données de plus de 400 sondages sur le bassin versant de Kervidy (Fig. II.11.a). Nous avons interpolé les valeurs d'épaisseur sur l'ensemble du bassin versant en utilisant une méthode de simulation séquentielle gaussienne (Johnson, 1987) adaptée et intégrée dans le package *Gstat* par Pebesma et Wesseling (1998). L'algorithme parcourt aléatoirement la carte et, pour chaque point visité dont il doit simuler la valeur, l'algorithme utilise un variogramme et les valeurs avoisinantes (du jeu de données de départ et des valeurs simulées précédemment) pour définir une distribution gaussienne locale et y tirer aléatoirement la valeur. Nous avons utilisé un modèle sphérique de variogramme pour approximer le variogramme expérimental tiré des sondages.

Cependant, cette méthode fait l'hypothèse que le jeu de données de départ suit une distribution normale, ce qui n'était pas le cas. Nous avons donc normalisé la distribution de données d'épaisseur des sols par une fonction de transfert par la méthode du Normal Score Transform (NST) proposée par Goovaerts (1997) avant de construire le variogramme expérimental et d'interpoler les valeurs d'épaisseur des sols. La distribution des valeurs simulées a été modifiée avec la fonction de transfert NST inverse. Pour réaliser le scénario alternatif « sols minces », nous avons procédé de même, mais en divisant par 2 les données d'épaisseur des sols. Un problème d'initialisation de *TNT2* sur cette distribution spatiale simulée nous a amené à la transformer par homothétie l'épaisseur minimale à 30 cm.

### **Spatialisation des types de sols**

*TNT2* permet de spatialiser par type de sols d'autres propriétés pédologiques relatives au comportement hydraulique des sols ou à des vitesses de transformation de la matière organique. L'étude pédologique des sols de Naizin (Walter et Curmi, 1998) nous a permis d'identifier trois grands types de sols en fonction de leur hydromorphie, et se distinguant par les valeurs de paramètres de *TNT2* suivantes (tableau II.1).

Le scénario alternatif « extension des sols hydromorphes » a été construit en utilisant l'indice topographique pente aval (Chaplot et *al.*, 2000) défini par l'équation 1 :

$$ITPA_i = \ln \left( \frac{\text{aire amont}_i}{\text{pente aval}_i} \right) \quad (1)$$

où  $ITPA_i$  est l'indice topographique pente aval au pixel  $i$  de la carte,  $\text{aire amont}_i$  est l'aire drainée par le pixel  $i$ , et  $\text{pente aval}_i$  est la pente moyenne du chemin hydrologique menant le pixel  $i$  au cours d'eau le plus



proche.

Cet indice prend des valeurs élevées dans les zones de fonds de vallées, caractérisés par de faibles pentes et des aires drainées fortes. Notre objectif a été de seuiller la carte de l'ITPA de façon à ce que l'union de cette surface et de celle actuellement cartographiée comme hydromorphe permette de doubler la surface occupées par des sols hydromorphes.

### **Structure verticale des sols dans TNT2**

Chaque pixel du bassin versant délimite une colonne où se superposent deux compartiments (Fig. II.18) :

- le sol : caractérisé par une porosité de rétention dans laquelle les racines peuvent puiser de l'eau. C'est également dans ce compartiment qu'on lieu les transformations de la matière organique (prélèvements par les cultures, minéralisation, dénitrification) ;
- le compartiment de drainage (sous-sol) : caractérisé par une porosité de drainage. Le niveau de la nappe dans cette colonne définit deux zones : la zone non saturée dans laquelle l'eau percole verticalement et la zone saturée dans laquelle la nappe s'écoule latéralement.

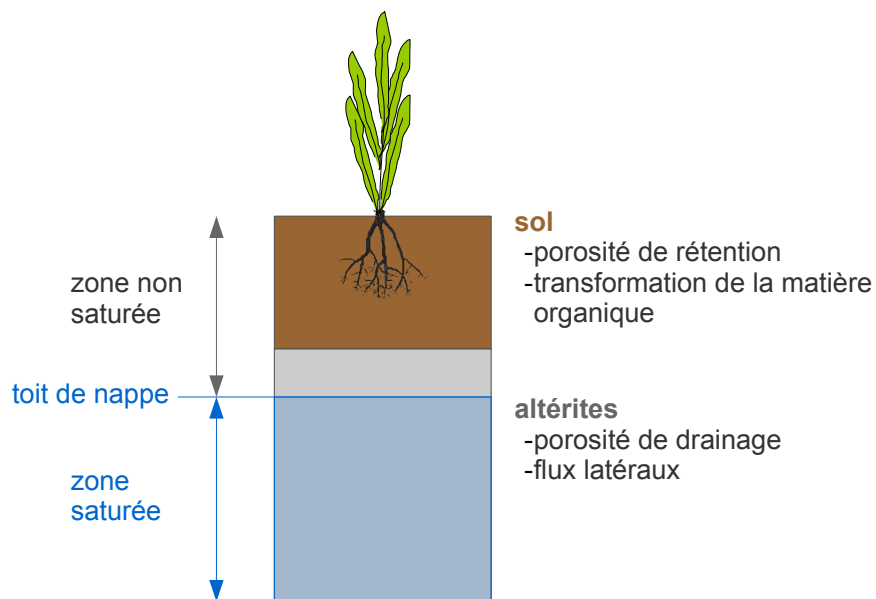


Fig. II.18 : Représentation schématique du milieu physique au sein d'un pixel de bassin versant dans TNT2.

Chaque compartiment est constitué de couches horizontales versant chacune dans celle du dessous afin de simuler le processus de percolation.

# Sources de la plateforme Qualscape

## ***Le projet Java***

La réalisation du projet a nécessité le développement d'un code informatique conséquent (environ 700 000 caractères de codes) relatif la création de paysages agricoles et l'analyse des sorties de simulation de *TNT2* et du modèle phosphore. Le langage Java a été choisi en raison de la disponibilité de bibliothèques existantes relatives à la manipulation de fichiers SIG, de bases de données MySQL et de fichiers XML. Enfin, l'ingénieur informatique de l'unité (Jordy Salmon-Monviola) a une bonne expérience de ce langage, avec lequel il a déjà développé quelques outils liés à des projets *TNT2*.

Le code source du projet Java est organisé en deux grands répertoires :

- */src* contient l'intégralité des sources du code informatique de *Qualscape* ;
- */use* contient des scripts d'utilisation du code informatique illustrant l'emploi et le maniement des classes de *Qualscape*.

Le code source des classes présentes dans */src* et ses sous-répertoires a été commenté en anglais au fur et à mesure du développement, afin d'expliquer pas-à-pas les algorithmes développés. Dans */use*, on trouvera des scripts relatifs à :

- la création et la manipulation de paysages agricoles : simuler l'évolution de l'occupation du sol (estivale et hivernale), exporter le paysage sous différents formats (notamment sous forme de base de données pour les outils relatifs à *TNT2*) ;
- l'analyse de paysages virtuels après couplage avec *TNT2* : bilan azoté annuel minéral et organique, couplage au modèle phosphore, visualisation et synthèse des sorties de *TNT2* sous formes d'indicateurs globaux (à l'échelle du bassin versant) et annuels.

À la racine du projet Java se trouve également un répertoire */resources* qui contient différentes données comme les matrices de transition de culture utilisées par Rotomatrix, différents arbres de décision (utilisés par Stochastree, le module d'interculture hivernale, le modèle phosphore).

## ***L'extension SpatialAnalysis.oxt***

Une extension au tableur d'OpenOffice.org (Calc) a été développée pour calculer les distances moyennes pondérées des cultures au corps de ferme par type d'exploitation. Le code source de l'extension a été également commenté en anglais pour expliquer les algorithmes développés. L'extension rajoute un bouton dans l'interface de Calc qui permet d'interroger une base de données MySQL pour effectuer les opérations et écrit les résultats directement dans le fichier Calc ouvert. Les informations identifiant la base de données interrogée sont à renseigner en tête du code source de l'extension.

Le fichier *SpatialAnalysis.oxt* permettant d'installer l'extension est disponible sur le wiki de mon doctorat<sup>31</sup>.

31 <http://147.99.166.141/wikidoctoratluc/doku.php?id=qualscape:spatialanalysis:accueil>

## ***Les scripts R***

Différents scripts ont été développés sous R pour construire les cartes d'épaisseur des sols (développements réalisés par Valérie Viaud) et pour effectuer les différents tests d'hypothèse relatifs au chapitre III. Les fichiers de scripts sont disponibles sur le wiki de mon doctorat, dans la rubrique de l'article relatif à l'article « Stochastic et occupation du sol »<sup>32</sup>.

---

32 <http://147.99.166.141/wikidoctoratluc/doku.php?id=articles:landusesimul:donnees:accueil>

## **Chapitre III**

*Stochastree*, un modèle de successions de cultures  
basé sur des arbres de décision stochastiques,  
intégrant des contraintes agronomiques



# Chapitre III : *Stochastree*, un modèle de successions de cultures basé sur des arbres de décision stochastiques, intégrant des contraintes agronomiques

## Résumé français

L'évaluation des impacts environnementaux des pratiques agricoles par des approches dynamiques repose sur la modélisation des processus et des flux à l'échelle des paysages, dont les temps de réponse peuvent s'étendre sur plusieurs décennies (Pauwels et al., 2001 ; Molénat and Gascuel-Oudou, 2002). De nombreux modèles ont ainsi été développés pour simuler la dynamique de l'eau et des nutriments, par exemple pour l'azote (Beaujouan et al., 2002 ; Grizzetti et al., 2005), le phosphore (Lewis and McGehan, 2002) et la matière organique (Gabrielle et al., 2002) ; leur mise en œuvre dans une optique d'évaluation environnementale suppose cependant l'existence de séries climatiques de longue durée et la description fine de l'occupation du sol et des pratiques agronomiques associées. L'occupation du sol doit être décrite à la fois avec une résolution spatiale suffisamment fine pour représenter le parcellaire et sur une étendue suffisamment grande pour décrire l'unité hydrologique du bassin versant et les interactions internes entre parcelles.

Pour faciliter la définition de scénarios agronomiques, l'objectif de ce chapitre est de développer un modèle de succession de cultures qui tienne compte de contraintes agronomiques, liées à des obligations techniques ou aux caractéristiques du milieu : objectifs de production par culture, distribution spatiale des cultures autour des sièges d'exploitation, allocation préférentielle des cultures selon l'hydromorphie des sols et la superficie de la parcelle. Nous avons développé un modèle nommé *Stochastree* basé sur des arbres de décision stochastiques, pour tenir compte du type d'exploitation et des caractéristiques des parcelles (superficie, distance au siège d'exploitation, culture actuelle, hydromorphie des sols) dans la procédure de simulation des successions de cultures. La démarche ne fait pas appel à une connaissance d'expertise préalable, mais repose sur un apprentissage par fouille de données sur un jeu de données réelles de successions de cultures sur plusieurs années.

Trois étapes successives ont été mises en œuvre : (1) l'identification des déterminants majeurs influençant la succession des cultures sur le bassin versant agricole de Naizin (Morbihan, France) d'une superficie de 1200 hectares. Cette analyse est faite pour l'ensemble du bassin versant puis par types d'exploitation (élevage laitier, élevage porcin, élevage mixte, autres exploitations) ; (2) la construction d'un modèle de transition de cultures fondé sur les arbres de décision stochastiques (*Stochastree*) et sa mise en œuvre pour simuler sur une période de 40 ans les successions des cultures d'été en tenant compte des contraintes agronomiques identifiées ; (3) la comparaison des performances respectives de *Stochastree* et d'un modèle de référence (nommé *Rotomatrix*), fondé sur un processus de Markov de premier ordre, pour restituer les contraintes agronomiques d'un jeu de validation.

Les arbres de décision de *Stochastree* et les matrices de transition de *Rotomatrix* ont été construits sur un même jeu d'apprentissage constitué des successions culturales observées sur le bassin versant de Naizin pour une période de 5 ans (1993-1997). Pour construire les arbres de décision manipulés par *Stochastree*, nous avons utilisé l'algorithme C4.5 (Quinlan, 1993) qui considère à la fois des variables discrètes (les types de culture, les classes d'hydromorphie des sols) et des variables continues (distance à la ferme, superficie des parcelles) sans hypothèse sur la distribution des variables. Chaque feuille de l'arbre de décision ne conduit pas à la prédiction d'un seul type de culture suivante, mais fournit les probabilités d'occurrence des différentes cultures suivantes. *Rotomatrix* implémente une chaîne de Markov de premier ordre temporel, qui estime une matrice de transition donnant pour chaque culture, sa probabilité de succéder à la culture précédente. Les arbres et les matrices ont été établis pour l'ensemble des parcelles du bassin versant, puis pour les sous-ensembles de parcelles appartenant à un même type d'exploitation.

La structure des arbres de décision de *Stochastree* s'est révélée proche des matrices de transition établies par *Rotomatrix*, la culture précédente apparaissant dans tous les arbres de décision comme le premier déterminant des successions culturales. Néanmoins, les arbres de décision construits pour les différents types d'exploitation agricole diffèrent par leur nombre de feuilles et l'ordre d'intégration des déterminants.

La performance des deux approches a été évaluée à partir d'un jeu de validation constitué des successions culturales observées sur une autre période de 5 ans (1998-2002) sur le même bassin versant, comparées aux simulations répétées 30 fois. *Stochastree* et *Rotomatrix* présentent des performances similaires dans la prédiction des successions de culture et le respect des objectifs de production. Par contre, *Stochastree* s'est révélé significativement supérieur pour le maintien de la distribution spatiale des cultures autour des sièges d'exploitation ainsi que pour l'allocation des cultures selon la classe d'hydromorphie des sols.

L'approche développée, fondée sur la construction d'arbres de décision, apparaît de ce fait facile à implémenter, explicite quant aux déterminants pris en compte et efficace pour respecter des contraintes agronomiques dans la simulation de successions de culture. Ces différentes qualités permettent d'envisager le couplage de *Stochastree* avec des modèles distribués de fonctionnement des bassins versants, comme par exemple des modèles de transfert des polluants ou des modèles de transmission de gènes.

Les matrices de transition et les arbres de décision stochastiques utilisés dans cet article sont présentés en Annexes II.

# Stochastree, a crop transition model based on stochastic decision trees, considering agronomic driving factors.

Luc Sorel<sup>a,b</sup>, Christian Walter<sup>a,b\*</sup>, Patrick Durand<sup>a,b</sup>, Valérie Viaud<sup>a,b</sup>;

<sup>a</sup> INRA, UMR1069, Sol Agro et hydrosystème Spatialisation, F-35000 Rennes

<sup>b</sup> Agrocampus Rennes, UMR1069, Sol Agro et hydrosystème Spatialisation, F-35000 Rennes

\* Corresponding author: christian.walter@agrocampus-rennes.fr INRA-Agrocampus Rennes, UMR SAS, 65 rue de Saint Briec, CS 84215, 35042 Rennes Cedex, France (phone: 0(+33) 2-23-48-54-39 / fax: 0(+33) 2-23-48-54-30)

Abbreviations: AWD, area-weighted distance; CTM, crop transition model; LCM, landcover change model; MCM, Markov chain model; SDT, stochastic decision tree; SRP, self-replacement probability.

Keywords: crop transition, decision tree, landcover change modeling, transition probability matrices, stochastic modeling.

## Abstract

Evaluating the environmental impacts of agricultural practices requires the use of simulation approaches to model hydrological fluxes at the landscape scale, whose response time can span over decades. The crop allocation over the field pattern determines the spatial distribution of fertilization practices, which are involved in soil and water quality issues. To facilitate the construction of agricultural scenarios, our objective was to develop a crop transition model able to account for agronomic and spatial driving factors: crop production objectives, spatial distribution of the crops around the farmsteads, and preferential allocation of crops on soil waterlogging classes. We developed an innovative model based on stochastic decision trees, *Stochastree*, to integrate farm-type and field characteristics (area, distance to farmstead, waterlogging, current crop) in the crop transition simulation process without prior expert knowledge, but relying on a data-mining approach. Simulation results of *Stochastree* were compared to field data and to a Markovian model based on transition matrices, *Rotomatrix*, to test its abilities to follow presumed agronomic constraints. Learned decision trees had a general structure similar to transition matrices. *Stochastree* and *Rotomatrix* exhibited similar performances in predicting crop transitions and outputting expected crop productions. *Stochastree* proved to be significantly superior in maintaining the spatial distribution of crops around the farmsteads and slightly better in allocating crops to the proper soil waterlogging class. The ease to construct decision trees suggests many potential couplings of *Stochastree* to various ecological models, like nutrient diffuse transfer models or gene transmission models.



## Introduction

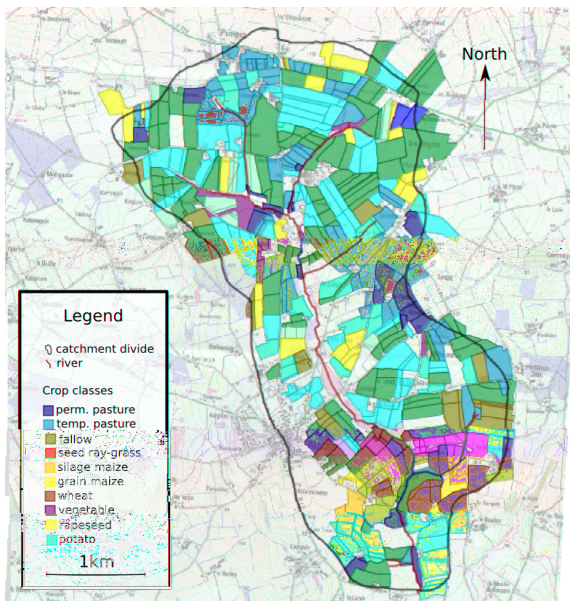
Despite their role in maintaining the agricultural potential of cultivated soils, organic and mineral fertilizers can also have adverse effects on water and soil qualities (Carpenter et al., 1998). Mismanagement of crop fields can lead to unwanted losses of nutrients via overland flow or vertical leaching. Such losses of nutrients alters the quality of the receiving freshwater systems and modify the expected soil nutrient budgets (Vadas et al., 2007). Several models have been designed to simulate the dynamics of water and nutrients – *eg* nitrogen (Beaujouan et al., 2002; Grizzetti et al., 2005), phosphorus (Lewis and McGehan, 2002), and organic matter (Gabrielle et al., 2002) –, but their use to assess the efficiency of best management practices (D’Arcy and Frost, 2001) requires large datasets containing complete climatic series of data and accurate spatio-temporal description of the farming landscape, including the crop management practices.

In the building of such datasets, the agricultural landcover must be described both at a fine resolution and over a large spatial extent in order to assess the efficiency of environmental remediation practices. The spatial resolution must be fine enough to represent the fields, which are the base units of the management of farm territories (Thenail and Baudry, 2004). Moreover, the relative positioning of the crops and their associated management practices influence the nutrient transfers to the groundwater and the river (Beaujouan et al., 2001). The spatial extent encompassed by the dataset must include the horizontal and vertical fluxes involved in non-point pollution, and the corresponding hydrological unit is the catchment (Beven, 2001). Concerning the duration covered by the dataset, landcover data must be available for a long period of time to account for (i) the hydrological time response of the catchment which can exceed several years (Pauwels et al., 2001; Molénat and Gascuel-Oudou, 2002), (ii) the dynamics of nutrient transformation (Gabrielle et al., 2002), particularly the residual effects of manure practices on soil phosphorus which can persist over decades (Russel, 1973; Karpinets et al., 2004).

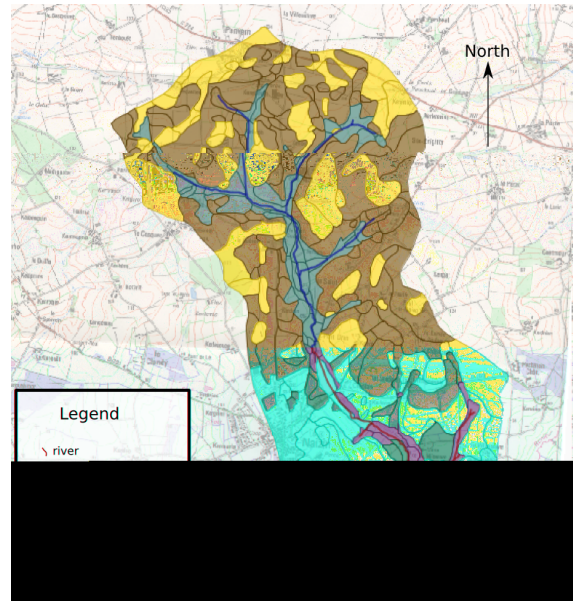
Agarwal et al. (2002) proposed a review of landcover change models based on a framework which categorizes the models by their manner to handle space and time dimensions, with human decision making processes. Considering the environmental impact assessment of agricultural practices, landcover models can be used (i) to extract state transition rules and sequences by mining datasets (Mignolet et al., 2004; Le Ber et al., 2006), (ii) to complete spatial data (Hanink and Cromley, 1998; Zhang and Li, 2008), (iii) temporally in a chronicle (Largouët and Cordier, 2001), or (iv) to simulate new states and evolutions (Baker, 1989; Walter et al., 2003), which can be optimized for crop productivity and the preservation of soil quality (Dogliotti et al., 2003; Rounsevell et al., 2003; Bachinger and Zander, 2007). Several approaches were developed to design distributed and multi-variate landcover change models (LCM), but the handling of both spatial and temporal constraints raises technical and scientific challenges (Verburg, 2006). Stratified Markov chains (Weaver and Perera, 2003; Ladet et al., 2005; Castellazi et al., 2007b; Pocewicz et al., 2008) require a large amount of data to compute the transition probabilities for each possible substrata (Usher, 1979; Baker, 1989) and hardly deal with spatial variability in the processes of landcover change (Huang et al., 2007). The calibration and validation of highly integrative models like cellular automata (Bockstael et al., 1995; Veldkamp and Fresco,

1996; Wu and Webster, 1998; Verburg et al., 2004), and explicitly rule-based models (Thornton and Jones, 1998; Largouët and Cordier, 2001; Gaucherel et al., 2006a) require arbitrary decisions and strong expertise (Li and Yeh, 2002; Houet and Hubert-Moy, 2006). Generalized linear models (Aspinall et al., 2004; Huang et al., 2007) rely on parametric hypotheses and are obscure to interpret, which complicates the design of new landcover change scenarios. On the contrary, decision trees (DT) are non-parametric methods known to be efficient without parameter tuning, accessible and useful for non-experts (Quinlan, 1981; Breiman et al., 1984; Murthy, 1998; Perlich et al., 2003). In the field of landcover modeling, DT are mainly used as classification tools based on remotely sensed data, pedological and geomorphological factors, or spatial descriptors of the farm territory (Lynn et al., 1995; DeFries and Chan, 2000; Friedl et al., 2002; Lawrence et al., 2004; Thenail and Baudry, 2004). Although DT have been used to simulate iterative temporal processes in other research fields (Gladwin, 1989; Heidenberger, 1996; Jordan et al., 1997; Hazen, 2002; Tarim et al., 2006), such applications of DT have not yet been done in landcover change modeling, to our knowledge.

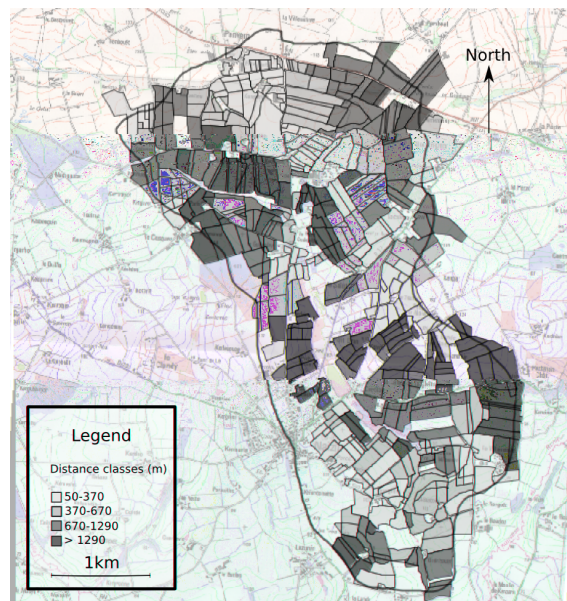
Our objectives were to bridge the gap between exploratory approaches and crop transition modeling by inferring stochastic decision trees trained by a data-mining analysis of a crop transition learning dataset, and to use them to simulate crop transitions over several decades with the incorporated agronomic driving factors. These objectives involved the following steps: (1) to identify relevant driving factors of crop transition on an agricultural catchment, (2) to build a crop transition model based on stochastic decision trees (named *Stochastree* hereafter) and simulate summer crop transition while accounting for agronomic driving factors, (3) to evaluate the abilities of *Stochastree* and a reference temporal first-order Markov chain model (named *Rotomatrix* hereafter) to comply with agronomic constraints. Finally, we discuss the structure of the models and the mutual benefits of a compared evaluation of *Stochastree* and *Rotomatrix* simulations.



(a)



(b)



(c)

Fig. III.19 : Agricultural land area of the Naizin catchment in 2000: (a) agricultural landcover of the fields associated to a known farmstead (adapted from Bordenave et al., 2005 and Payraudeau et al., 2006); (b) spatial distribution of soil waterlogging classes (adapted from Walter and Curmi, 1998); (c) direct euclidean distance between a field and its farmstead (distance classes are quartiles).

## Materials and methods

### *Presentation of the Naizin catchment*

The Naizin catchment has been extensively studied since 1971 by various French research organizations (Cheverry, 1998). It is a 15 km<sup>2</sup> agricultural catchment located in the south of Brittany (France), 400 km west of Paris. Three datasets were combined here (Fig. III.19): a landcover survey of summer crops done on-field between 1993 and 2002 by Bordenave et al., (2005), a life-cycle analysis of farmstead practices done in 2003 by Payraudeau et al., (2006), and a soil map done by Walter and Curmi (1998). The landuse of Naizin is dedicated mainly to agriculture (12 km<sup>2</sup>). The agricultural landcover survey is remarkably exhaustive and the dynamics of crop areas suggests a relative stationarity in crop transition both at the catchment scale (Fig. III.20) and for the different farm-types. The original agricultural landcover data was homogenized and aggregated into the following crop classes to keep the information consistent through time:

- permanent pasture;
- temporary pasture;
- fallow;
- seed ray-grass;
- vegetable (mainly beans);
- rapeseed;
- grain maize and silage maize. In 2001 and 2002, grain and silage maize were not distinguished in the survey but observed as maize. Therefore grain and silage maize data are not available for these two years;
- winter wheat (this class also includes other winter cereals like barley);
- potato.

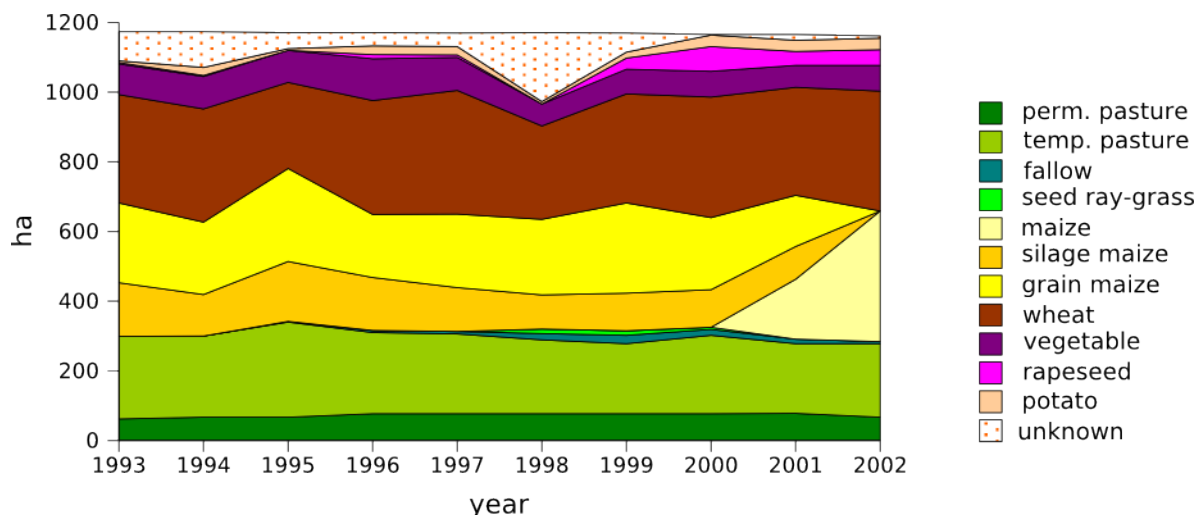


Fig. III.20: Dynamics of the cultivated areas (ha) in the Naizin catchment, 1993-2002, resulting from annual inventory.

The life-cycle analysis survey conducted in 2003 with the farmers (Payraudeau et al., 2006) located the farmsteads of the catchment, and depicted a typology of the farming activities composed of four farm-types:

- dairy (13 farmsteads), which concerned 33% of the total agricultural area;
- pig (6 farmsteads), 34%;
- dairy-pig (10 farmsteads), 19%;
- other (8 farmsteads), 14%, which artificially grouped (i) farmsteads of unknown activity, and (ii) farmsteads with marginal farm activities.

Given the locations of the farmsteads (Payraudeau et al., 2006) and of their fields (Bordenave et al., 2005), we associated a distance-to-farmstead to each field. The euclidean distance between the farmstead and the centroid of the field polygons seemed a reasonable estimator of the accessibility of the fields because the road network is rather dense and isotropic in this catchment. Fig. III.19.c shows the distance-to-farmstead of the field in year 2000 and the road network of Naizin.

The nine original classes of soil waterlogging described in the soil map of the catchment (Walter and Curmi, 1998) were aggregated into the following classes (Fig. III.19.b):

- 1 (well-drained): no soil redoximorphic feature within the first 80 cm. This class represents 43% of the total agricultural area;
- 2 (intermediate): weak soil redoximorphic features at less than 80 cm, 44%;
- 3 (waterlogged): intense redoximorphic features from the surface, 13%.

Each field was characterized by the dominant soil hydromorphy class represented within its limits. This waterlogging attribute can be considered constant through time since most of the artificial drainage systems were setup before 1993 in this catchment.

The topography of the catchment is relatively smooth (mean slope is below 5%, with maximal values of 14%), its highest point is at 136 m above the sea level while its outlet is at 60 m (Walter and Curmi, 1998). The overall gentle slopes are not limiting for the preparation and cultivation of crops.

### ***Agronomic constraints on the agricultural landuse of Naizin***

Beside the crop succession itself, which is known to play a key role in the farm management of soil fertility and nutrients dynamics (Thornton and Jones, 1998), we tested the relevance of four presumed spatial driving factors of crop transition in Naizin:

- *farm-type*: farmstead sharing the same activity and main production (milk, pigs, etc.) are submitted to the similar constraints (Thenail and Baudry, 2004). Moreover, the life-cycle analysis also revealed that fields of some farmsteads of Naizin were outside the surveyed area, and similarly, that some surveyed fields of Naizin belonged to farmsteads located just outside the catchment divides. The existence of these sparse field patterns made it more relevant to study the agricultural driving factors of crop transition at the farm-type scale instead of the farm scale;

- *crop production objectives*: the allocation of crops on the field pattern of farm fulfills production objectives set by the farmer (Thenail and Baudry, 1994), thus farm-types may be characterized by specific crop production objectives. This is particularly true for fodder production (grass and silage maize), supposed to fulfill the need of the cattle; these production objectives will be assessed using the area dedicated to each crop, assuming that the yield potentials are homogeneous over the whole catchment;
- *spatial distribution of crops*: to optimize the use of the farm labor time, fields implying a lot of operations (soil work, field input and crop management, shifting or milking of cattle) are located closer to the farmstead (Baudry et al. 2006). Farm-types may be characterized by specific mean distances between the crop fields and the farmstead;
- *soil waterlogging*: some crops are sensitive to soil waterlogging (Dogliotti et al., 2003; Houet and Hubert-Moy, 2006).

Topography was not included as a significant driving factor because of the observed gentle slopes. We assumed that these spatial driving factors were time-invariant over the 1993-2002 period because previous studies showed that a combination of driving factors can be steady over several decades (Aspinall, 2004), particularly when socio-economic factors are stable over the observed period (Pocewicz et al., 2008), which was the case. Moreover, between 1993 and 2002, the area exchanged between farmsteads was less than 50 ha and concerned 8 of 37 farmsteads, without significantly modifying their field pattern. The significance of the presumed spatial factors was tested by non-parametric Wilcoxon comparison tests ( $p=0.95$ ) between a set of observed annual indicators with either a target value or another of indicators.

### **Crop production objectives**

To test the existence of a farm-type effect on crop production objectives, we compared the crop proportions of a farm-type (expressed as percentages of the area occupied by the farm-type) to the proportions observed on the whole catchment area.

### **Spatial distribution of crop fields**

To test the existence of spatial structures specific to crops in the field pattern of the farmstead, we calculated the mean distance weighted by the field area (area-weighted distance: AWD) as follows:

$$AWD(ft, c, y) = \frac{\sum_{field=1}^{F(ft, c, y)} distance_{field} \times area_{field}}{\sum_{field=1}^{F(ft, c, y)} area_{field}} \quad (1)$$

where  $AWD(ft, c, y)$  is the AWD (distance in m) calculated with fields belonging to farm-type  $ft$ , occupied by crop  $c$  at year  $y$ ;  $F(ft, c, y)$  is the ensemble of such fields and  $field$  is one of its elements characterized by its distance-to-farmstead  $distance_{field}$  (in m) and its area  $area_{field}$  (in  $m^2$ ). This integrated indicator of the crop spatial distribution allowed us to test two hypotheses:

- farm-types have specific spatial structures: does the crop AWDs computed for a given farm-type significantly differ from the crop AWDs computed over the whole catchment?
- within a farm-type, crops have specific spatial structure: does the farm-type crop AWDs significantly differs from the global farm-type AWD?

### **Preferential allocation of crops with soil waterlogging**

13% of the total agricultural area is located on waterlogged soil. For each crop, this value was compared with the series of annual crop proportions located on this class (expressed as a percentage of the total crop area).

### **Structure of the crop transition models**

For comparison purposes, we implemented two stochastic models whose purpose was to simulate the summer crop transition at the field level. *Rotomatrix* is an implementation of a temporal first order-Markov chain and will serve as a reference model for comparison with *Stochastree*, which is based on stochastic decision trees. Both model types were implemented either with consideration for farm-types (hereafter called specific models) or without (hereafter called generic models).

### **Rotomatrix, a temporal first order Markov chain**

Markov chains have been extensively used both in landuse/landcover change and crop transition modeling, therefore *Rotomatrix* description is only succinctly described here. More details concerning Markov chain formalism in the landcover change context can be found in Anderson and Goodman (1957), Usher (1992), Logofet and Lesnaya (2000), Coppedge et al. (2007), and Castellazzi et al. (2007b). Our approach of crop transition modeling followed a discrete description of time (annual summer crop transition) and a finite number of crop classes (the 10 crop classes observed in Naizin during 1993-2002). It is therefore possible to represent mathematically all the possible crop transitions as a temporal first order Markov chain with a square matrix  $P$  of dimension  $m$  (where  $m$  is the number of crop classes) in which each  $P_{ij}$  represents the probability for a crop  $j$  to succeed after the previously grown crop  $i$ . The  $P$  matrix can be built experimentally from a dataset by (i) counting all the observed annual transitions from a crop  $i$  to a crop  $j$ , (ii) converting these frequencies of transition into probabilities as follows:

$$P_{ij} = \frac{n_{ij}}{\sum_{j=1}^m n_{ij}} \quad (2)$$

where  $n_{ij}$  is the number of transitions from crop  $i$  to crop  $j$  and  $m$  is the number of crop classes.

A temporal first order Markov chain model particularly assumes the following properties:

- the transition probabilities to the succeeding crop  $j$  are not statistically independent of the possible previous crop  $i$  so that they could possibly form a first order Markov chain. The statistical independence of the probability matrix  $P$  is tested by the following statistic:

$$\alpha = 2 \sum_{i=1}^m \sum_{j=1}^m n_{ij} \ln \left( \frac{P_{ij}}{P_j} \right) \text{ with } P_j = \sum_{i=1}^m \left( \frac{n_{ij}}{\sum_{i=1}^m \sum_{j=1}^m n_{ij}} \right) \quad (3)$$

where  $\ln$  is the natural logarithm,  $P_{ij}$  is the probability of transition from crop  $i$  to crop  $j$ ,  $P_j$  is the marginal probability for the  $j^{\text{th}}$  column,  $n_{ij}$  is the number of transitions from crop  $i$  to crop  $j$ , and  $\alpha$  is distributed asymptotically as  $\chi^2$  with  $(m-1)^2$  degrees of freedom. The null hypothesis that annual landcover transitions are statistically independent (*ie*, totally random) is rejected with a 0.95 confidence (against the alternative hypothesis that the transitions are not independent and that they could form a first order Markov chain) if:

$$\alpha > \chi_{p=0.05, df=(m-1)^2}^2 \quad (4)$$

- the probabilities are stationary in time: the  $P_{ij}$  are time independent. Often assumed for exploratory purposes (Usher, 1992; Baker, 1989; Logofet and Lesnaya, 2000) the time-independence of the transition probabilities can be theoretically tested with the following statistic:

$$\beta = 2 \sum_{i=1}^m \sum_{j=1}^m \sum_{t=1}^T n_{ij}(t) \ln \left( \frac{P_{ij}(t)}{P_{ij}} \right) \quad (5)$$

where  $\beta$  is a derivation of Eq (3),  $P_{ij}(t)$  and  $n_{ij}(t)$  are respectively the probability and the frequency of transition from crop  $i$  to crop  $j$ ,  $T$  is the number of time intervals. The null hypothesis that  $P_{ij}(t)$  is constant over time is rejected with a 0.95 confidence (against the alternative hypothesis that the transitions are not time-independent) if:

$$\beta > \chi_{p=0.05, df=m(m-1)T}^2 \quad (6)$$

- the transition probabilities between time steps  $m$  and  $n$  can be computed with an intermediary time step  $t$  by the multiplication of the following matrices (the Chapman-Kolmogorov equation):

$$P(m, n) = P(m, t) P(t, n) \quad (7)$$

where  $P(m, n)$  is the transition probability matrix computed with Eq (2) and  $n_{ij}(m, n)$  is the number of fields with crop  $i$  at time  $m$  and crop  $j$  at time  $n$ .

*Rotomatrix* can handle several transition probability matrices on a catchment field pattern, each matrix corresponding to a farm-type set of probabilities.

### **Stochastree, a stochastic model based on probability estimates of decision tree leaves**

As formalized by Quinlan (1981) and Breiman et al. (1984), the purpose of a DT is to answer a specific question through the means of an organized succession of *nodes* where boolean tests occur, which finally lead to a decision *leaf* which outputs the answer to the question. DTs are mainly used to classify *instances* carrying a set of *attributes*. In the context of this paper, the question addressed by the decision tree is ‘‘What is the next crop?’’ and an instance is a field of the landscape characterized by the following attributes: its current landcover, its area, its distance-to-farmstead, its soil waterlogging class. Each node tests the *value of an attribute* and depending on the issue of the test, branches the tested instance to other nodes until a leaf is



reached. The graphical structure of DTs makes them easy to read, to interpret or to design, in comparison to black-box classifiers like neural networks, generalized linear models or maximal likelihood classifiers (Pal and Mather, 2003). Such DTs are deterministic, so that two instances carrying the same values for their attributes will be routed by the same test nodes to identical decision leaf.

DT do not require to be built with expert knowledge and they can be extracted from a learning dataset. We used the C4.5 algorithm (Quinlan, 1993) because it can handle both discrete (the 10 crop classes, the three soil waterlogging classes) and continuous attributes (distance-to-farmstead and field area are real numbers) without making any assumption on the distribution of the input attributes (Friedl and Brodley, 1997). The learning dataset supplied to C4.5 must contain instances characterized by *explanatory attributes* (the observed current crop, field area, field distance-to-farmstead, soil waterlogging class) and one *target attribute* (the observed next crop). The learning process starts with the “tree growing” phase, which recursively subdivides the learning dataset into smaller partitions, by testing the values of explanatory attributes and maximizing the information gain ratio. The information carried by a dataset partition is evaluated by an indicator derived from the Shannon entropy index (Shannon, 1948):

$$info(T) = - \sum_{j=1}^m \left( p_{j/T} \log_2(p_{j/T}) \right), \quad \text{with } p_{j/T} = \frac{n_{j/T}}{\sum_{j=1}^m n_{j/T}} \quad (8)$$

where  $info(T)$  is the information entropy of data partition  $T$ ,  $\log_2$  is the logarithm function to base 2,  $m$  is the number of crop classes,  $p_{j/T}$  is the proportion of instances in  $T$  carrying the crop class  $j$ .  $P_{j/T}$  is therefore the ratio between  $n_{j/T}$ , the number of instances carrying class  $j$  in  $T$ , and the total number of instances of  $T$ .

Given a test  $X$  that partitions  $T$  into  $n$  outcomes, the total information content after applying  $X$  is the sum of the information of the sub-partitions weighted by the number of instances in each sub-partition:

$$info_X(T) = \sum_{i=1}^n \left( \frac{\sum_{j=1}^m n_{j/T_i}}{\sum_{j=1}^m n_{j/T}} info(T_i) \right) \quad (9)$$

The information gained by splitting  $T$  using  $X$  is:

$$gain(X) = info(T) - info_X(T) \quad (10)$$

The gain criteria selects the test for which  $gain(X)$  is maximum.

To compensate for favoring tests with large number of splits,  $gain(X)$  is normalized by:

$$split\ info(X) = - \sum_{i=1}^n \left( \frac{\sum_{j=1}^m n_{j/T_i}}{\sum_{j=1}^m n_{j/T}} \right) \log_2 \left( \frac{\sum_{j=1}^m n_{j/T_i}}{\sum_{j=1}^m n_{j/T}} \right) \quad (11)$$

Therefore the final splitting metric is:

$$gain\ ratio(X) = \frac{gain(X)}{split\ info(X)} \quad (12)$$

A better information gain ratio thus indicates that the instances in the new partitions are more homogeneous concerning the targeted “next crop” value. However, a noisy learning dataset artificially generates unnecessary test nodes, causes overfitting, and alters the robustness of the DT. Therefore, most DT algorithms use a stop rule (like “stop subdividing a partition if it contains less than  $n$  instances”) and even a “pruning phase” during which subtrees not contributing significantly towards generalization accuracy are replaced by a final leaf (Murthy et al., 1998). Such pruned DT are usually more robust although their final leaves are not perfect (they contain different values for the target attribute). At *imperfect leaves*, the algorithm chooses the most represented value of the target attribute and sets it as the leaf decision.

The originality of our approach was (i) to expand the imperfect leaves of the deterministic trees into *stochastic nodes* by converting the frequencies of the target values into probabilities estimates (Quinlan, 1990; Provost and Domingos, 2003; Liang et al., 2006), (ii) and to use them as transition probabilities to simulate the next crop. The deterministic trees were built using the C4.5 algorithm implemented in the Weka data-mining software (Witten and Frank, 2005) with the following parameters:

- tree growth parameters: 10-fold cross-validation, growth should stops at a node containing less than five instances (parameter set by default to two, but raised to five to avoid noise artifacts);
- tree pruning parameters: three of the 10 folds were randomly selected and used for pruning; pruning confidence rate was 25% (recommended default value); sub-tree raising and reduced error mode were activated.

The deterministic trees were then expanded into stochastic decision trees and saved as XML files, whose the tree-structured syntax is particularly well adapted to DT representation and parsing. A program, *Stochastree*, was designed on purpose. It handles landcover dataset and parses the XML decision trees. When a stochastic node is met, a random number is drawn in  $[0, 1[$  (uniform law) and is compared to the cumulated probabilities of the possible outcomes of the node (which sum up to 1). The outcome corresponding to the interval in which the random number falls is chosen.

## ***Comparative analysis of the simulations***

### **Prediction accuracy for annual crop transition**

The construction of both *Rotomatrix* and *Stochastree* models requires a learning dataset to build transition matrices and decision trees. Since we wanted to test the model abilities to predict crop transition on a validation dataset, the Naizin 1993-2002 dataset was arbitrarily divided in two: the learning dataset contained 2409 instances derived from the annual crop transitions observed in Naizin during the 1993-1998 period, while the validation dataset contained the 1799 crop transitions that occurred during the 1998-2002 period. For each crop transition of the validation dataset, 30 transitions were simulated by each model as statistical repetitions. Then, contingency tables between the expected and the simulated crop provided a way to compare the prediction efficiency of *Rotomatrix* and *Stochastree* with a pure random classifier.

### **Model abilities to follow agronomic constraints**

Once trained on the learning dataset, generic and specific variations of *Rotomatrix* and *Stochastree* model were used to simulate crop transition during 40 years from the status of Naizin on year 2000, which was the most complete year in terms of agricultural landcover observations. For each resulting simulation, the following statistical tests to evaluate the model ability to comply with agronomic constraints by comparing the 40-year simulation results (2000-2040) with the observed dataset decade (1993-2002) were done with Wilcoxon distribution tests ( $p=0.05$ ):

- crop production objectives of the farm-types: for each farm-type, the simulated area proportions of crop were compared to the observed proportions;
- spatial distribution of the crop: the simulated farm-type crop AWDs were compared to the observed AWD;
- preferential allocation of crops related to soil waterlogging of fields: simulated percentages of each crop allocated to the waterlogged soil class was compared with observed proportions.



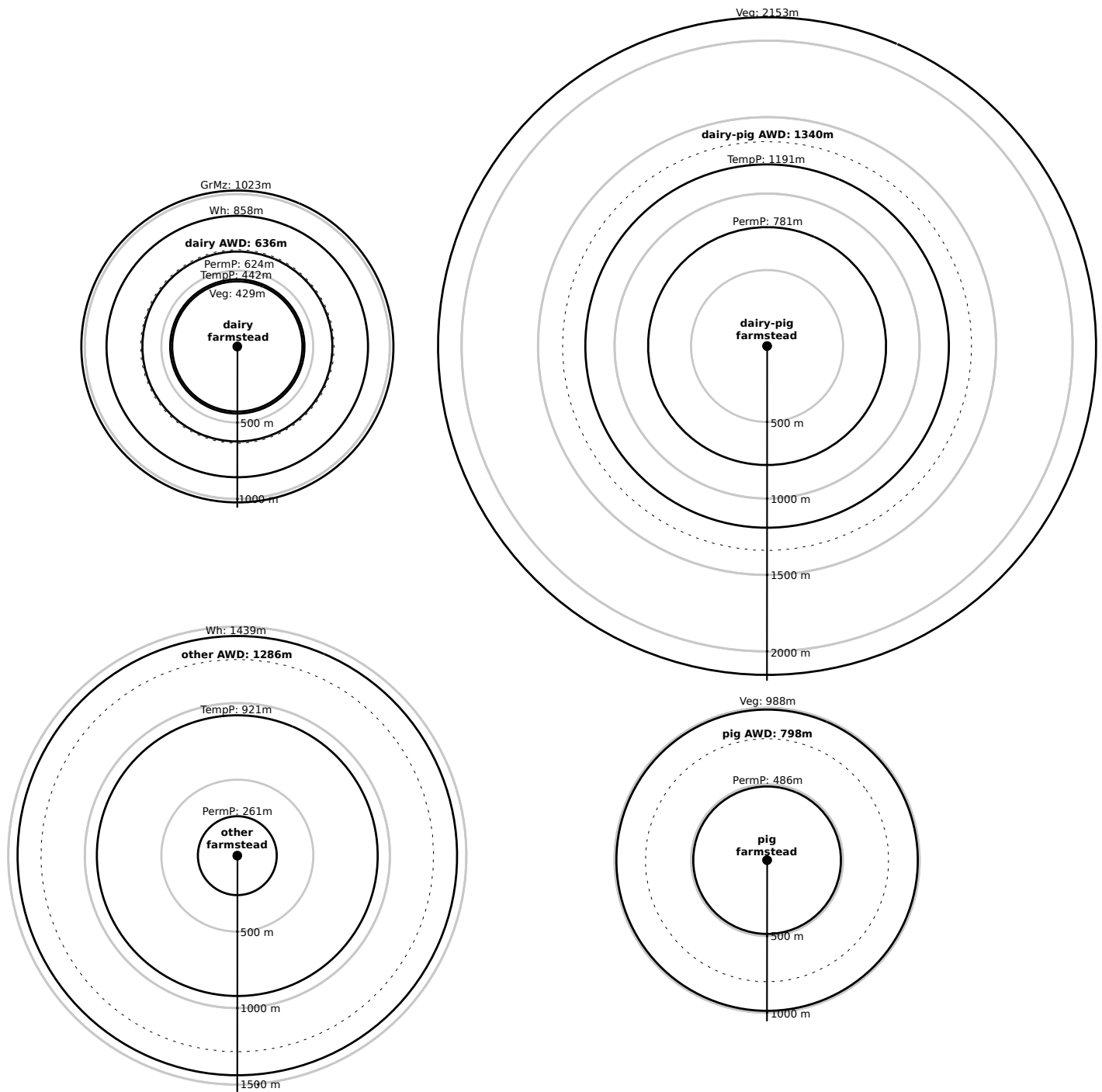


Fig. III.21: Spatial distribution of the crops observed in Naizin (1993-2002) around the four farm-types (as if viewed from the sky): dairy, dairy-pig, pig, and other. The circles show the mean area-weighted distance (AWD) expressed in meters. Only the crops for which the AWD were significantly different than the global AWD of the farm-type are shown. Crop classes are: Fal=fallow, TempP=temporary pasture, PermP=permanent pasture, SRG=seed ray-grass, Wh=wheat, GrMz=grain maize, SIMz=silage maize, Pot=potato, Rape=rape, Veg=vegetable.

## Results

### ***Driving factors of crop transition in Naizin***

Results of the Wilcoxon tests performed on observed data of Naizin exhibited crop production objectives specific to each farm-type. The dairy farm-type was characterized by significantly larger proportions for fodder crops (silage maize, permanent and temporary pastures) and smaller proportions for wheat, grain maize and vegetables. On the contrary, pig farm-type was characterized by smaller proportions for fodder crops; and larger proportions for wheat, grain maize, and potato. The dairy-pig farm-type exhibited intermediate proportions between those of the dairy and the pig farm-types.

The AWD spatial indicator revealed two levels of spatial structures in the crop pattern of Naizin:

- most farm-type crop AWDs were significantly different than the crop AWDs computed over the whole catchment (except marginal crops like fallow, rapeseed, and potato), thus showing a farm-type effect on crop spatial structures;
- the comparison of specific crop AWD (*i.e.* calculated for a given farm-type) to the global farm-type AWD showed that some crops were significantly close or far away from the farmstead (Fig. III.21). Dairy farm-type exhibited the most spatially-structured crop pattern, and the crops with significant spatial structures were permanent and temporary pastures (closer to farmstead), wheat, grain maize and vegetables (farther to farmstead).

Concerning soil waterlogging, fallows and permanent pastures proved to be preferentially allocated on waterlogged soils (respectively 51% and 40% of each crop area), whereas all the other crops were preferentially allocated on well-drained soils.

### ***Structure of the crop transition models***

#### ***Rotomatrix models***

We tested the statistical independence of the *generic* and the four *specific* transition matrices constructed from the learning dataset (1993-1998). The dimension of these five matrices ( $m$ ) was 10 since 10 crop classes were represented on the catchment. The values of  $\alpha$  for the generic and farm-type specific ranged between 546 and 4767, largely exceeding the critical value of 103 ( $\chi^2_{p=0.05, df=9 \times 9}$ ). Therefore, the annual crop transitions are not random and can be modeled with at least a first order Markov chain. Generic and specific transition matrices were all ergodic (the first eigenvalue equaled 1, others were strictly lower than 1) and the temporal stationarity of the Markovian processes can thus be assumed, and visually controlled (Fig. III.20) although we did not test it. Following the methodology of Coppedge et al. (2007), we used  $\chi^2$  tests to control that matrices were significantly different among farm-types.

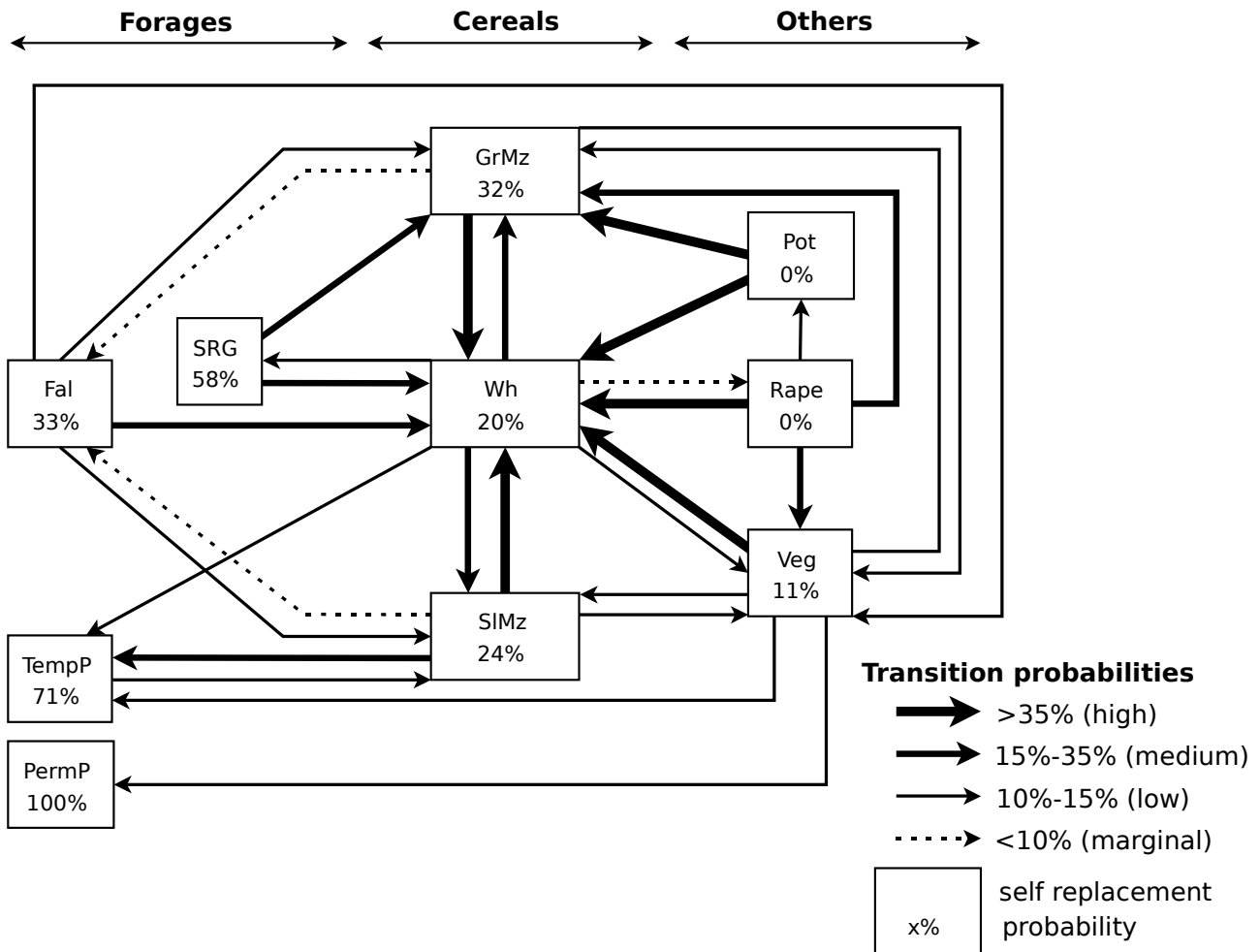


Fig. III.22: Box and arrow diagram of the generic matrix illustrating crop transitions for Naizin catchment, 1993-1998. For clarity, only transition probabilities >10% were included, except for crops generated only by a marginal transition probabilities (SRG, Rape, Fal, and PermP). Values within boxes are self-replacement probabilities whereas the arrows indicates transition probabilities. Landcover classes are: SRG=seed ray-grass, Fal=fallow, TempP=temporary pasture, PermP=permanent pasture, Wh=wheat, GrMz=grain maize, SIMz=silage maize, Pot=potato, Rape=rapeseed, Veg=vegetable.

The generic transition matrix is represented in Fig. III.22 as a box-and-arrow diagram (Coppedge et al., 2007). Cereals, and particularly wheat, played a major role in crop transitions since they were pointed by the highest transition probabilities (wheat was the only crop targeted by probabilities superior than 35%). The dynamics of cultivated areas on Naizin (Fig. III.20) showed that fallow, potato, rapeseed and seed ray-grass were marginal and intermittent crops. This was modeled with marginal or low probabilities pointing to them (meaning that other crops would seldom turn into them), and high probabilities to turn into a different crop the following year. Permanent pastures have a 100% self-replacement probability and a marginal probability originated from vegetables pointing to it, which means that the area occupied permanent pastures can only increase with time. Such a crop in a Markov chain is called an “absorbing state” (Usher, 1979).

The probabilities of specific transition matrices differed mainly for forages (targeted with higher probabilities for the dairy farm-type, lower probabilities for the pig farm-type) and secondarily for vegetables (targeted with higher probabilities for the pig farm-type).

## Stochastree models

For the sake of clarity, Fig. III.23.a represents only the general structure of the generic and farm-type specific stochastic decision trees (SDT). The root test node of all the SDT involved the “current crop” attribute, which makes them very similar to a temporal 1<sup>st</sup> order Markov chain. But the intermediary tests between the “current crop” root test and the “next crop” decision leaves stratify the transition probabilities by waterlogging classes, area and distance splits. However, more than a half of the decision branches resulting from the “current crop” root test lead directly to a stochastic “next crop” leaf, which therefore carry the same probability distribution than the corresponding transition matrix row since *Rotomatrix* and *Stochastree* models were trained on the same learning dataset. Particularly, the SDT of the least represented farm-type (dairy-pig) contained only one root branch implying other explanatory attributes (“if current crop is wheat”). Depending on the farm-type, the “current crop is wheat” branch involved subtrees with different combinations of tested attributes.

(a)

(b)

Simplified structure of the stochastic decision trees (SDT)

A sample of the dairy SDT: the branch for a TempP current crop

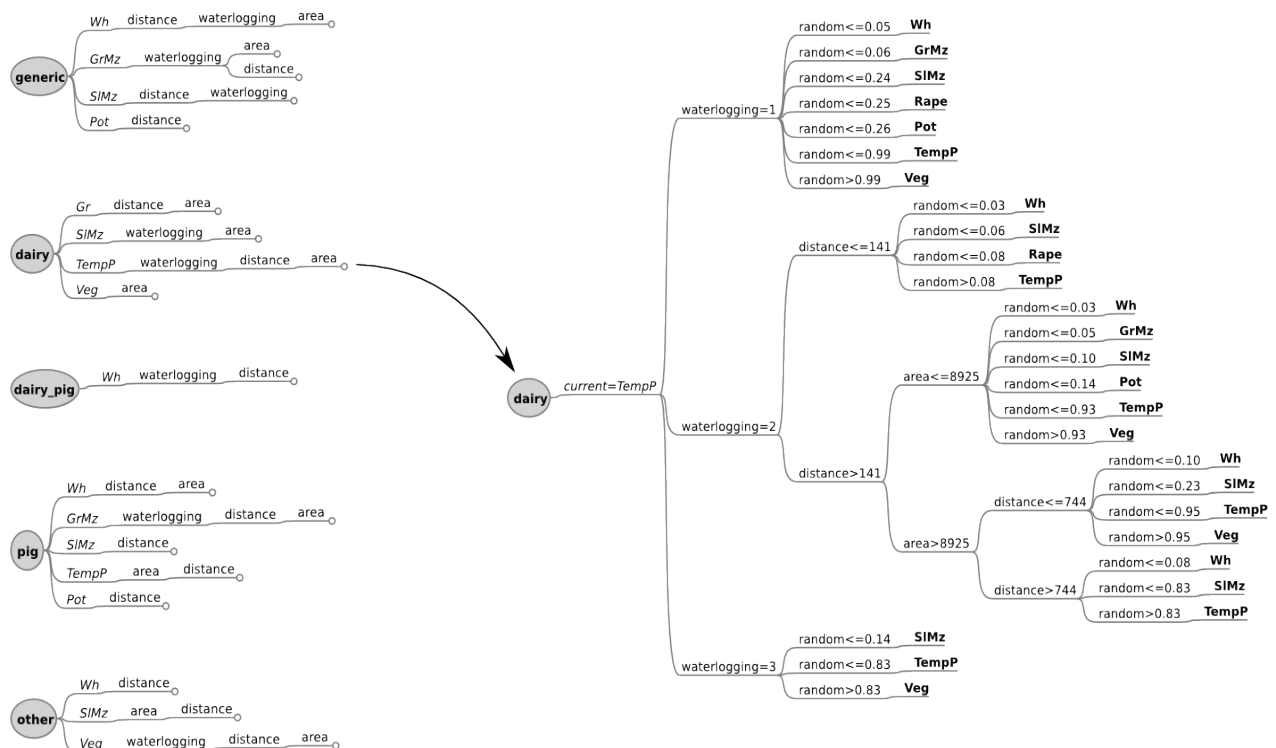


Fig. III.23: Stochastic decision trees (SDT) used to simulate the landcover transitions. SDT should be read from left to right. Root tests are in italics (all SDT started by testing the current crop), decision leaves are in bold.

- (a) The simplified structures only represent the succession of attributes tested by the DT for the branches where other attributes than the current crop appeared;
- (b) A sample of the dairy SDT is shown, corresponding to the TempP current crop. Distances and areas are expressed in m and m<sup>2</sup> respectively. Waterlogging classes are: 1=well-drained, 2=intermediate, and 3=waterlogged. Crop classes are: PermP=permanent pasture, TempP=temporary pasture, Fal=fallow, SRG=seed ray-grass, SIMz=silage maize, GrMz=grain maize, Wh=wheat, Veg=vegetable, Rape=rape, Pot=potato.



expected crop	simulation method				
	Rotomatrix		Stochastree		random
	generic	specific	generic	specific	
<i>permanent pasture</i>	97	97	97	97	10
<i>temporary pasture</i>	52	54	52	55	10
<i>fallow</i>	21	31	22	32	10
<i>seed ray-grass</i>	63	54	63	59	11
<i>silage maize</i>	16	18	14	17	9
<i>grain maize</i>	30	32	34	34	11
<i>wheat</i>	36	39	38	41	9
<i>vegetable</i>	13	10	12	12	10
<i>rapeseed</i>	0	0	1	0	10
<i>potato</i>	3	5	4	7	10
<i>(all)</i>	38	40	39	41	10

Table III.2: Success rate of the crop transition models by expected crop classes on a validation dataset. Compared models are Rotomatrix, Stochastree, and the pure random predictor (each one of the 10 crop classes was given a 0.1 transition probability).

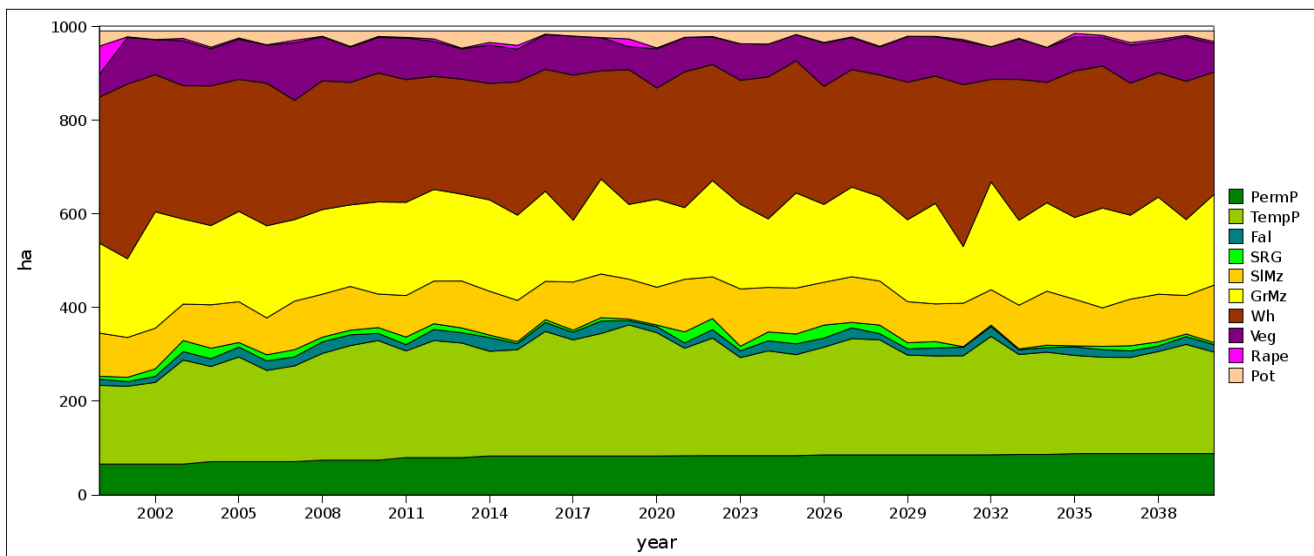


Fig. III.24: Dynamics of the cultivated areas (ha) in Naizin simulated during 40 years by Stochastic specific decision trees, starting from the observed landcover in 2000. Crop classes are: PermP=permanent pasture, TempP=temporary pasture, Fal=fallow, SRG=seed ray-grass, GrMz=grain maize, SIMz=silage maize, Wh=wheat, Veg=vegetable, Rape=rape, Pot=potato

The sample branch from the dairy SDT (Fig. III.23.b) illustrates how various attributes can be successively tested and lead to different sets of transition probabilities. It shows that once a test node has partitioned the tree by testing a categorical variable (like waterlogging) the subtrees corresponding to each possible waterlogging class are pure concerning this attribute, thus the attribute will not be tested further down the tree structure. On the contrary, numeric attributes (like area and distance) may be tested at any stages of the decision tree since test nodes perform binary splits of the dataset on them.

## **Comparative analysis of the simulations**

### **Accuracy to predict expected annual crop transitions on a validation dataset**

Table III.2 shows the ability of the *Rotomatrix* and *Stochastree* models to predict the next crop on a validation dataset representing all the annual landcover transitions observed in Naizin between 1998 and 2002. However, the purpose of stochastic models is not to predict accurately the most probable expected landcover, but to recreate the variability observed in the learning dataset. Therefore, the simulations were compared with the performances of a pure random classifier. The overall prediction rate of the crop transition models was around 40% significantly better than the pure random classifier for every crop (overall prediction rate of 10%), excepted for the most marginal crops (vegetable, rapeseed and potato). Successful prediction rates of specific models for non-marginal crops were above success rates of generic models by 2 to 10%. *Stochastree* performed slightly better than *Rotomatrix* (except for silage maize and vegetable) with a difference of 1% for the overall prediction rate.

### **Model ability to comply with agronomic constraints**

Each generic and specific variation of *Rotomatrix* and *Stochastree* models was used to simulate summer crop transition for a duration of 40 years, starting from the status of Naizin in 2000. Simulations resulting from the four models were similar on the dynamics of cultivated areas at the catchment scale (Fig. III.24 illustrates the resulting dynamics of the cultivated areas simulated by specific SDT): (i) a very slow increase of permanent pastures area; (ii) a rapid but small augmentation of the area dedicated to temporary pastures the first 10 years; and (iii) a general stationarity of the crop areas. These 40-year simulated datasets were compared to the observed 1993-2002 dataset to test their abilities to comply with agronomic constraints.

### **Crop production objectives of the farm-types**

The resulting simulated crop areas were expressed in percentages of the farm-type total area and compared with production objectives observed on the 1993-2002 period. Table III.3 shows that the simulations resulting from *Stochastree* and *Rotomatrix* modified significantly the simulated crop proportions compared to observed proportions, particularly for the pig and dairy farm-types, which had specific very specific objectives for wheat, maize, and pastures. Generic transition models implied larger differences of proportions than specific transition models for most crops. The significant differences of proportion implied by *Stochastree* were generally higher by 2-5% than the proportion differences implied by *Rotomatrix*. The sign

of significant differences was generally the same among transition models for a given crop.

farm-type	dairy				dairy-pig				pig			
	<i>Roto gen</i>	<i>Roto spec</i>	<i>Sto gen</i>	<i>Sto spec</i>	<i>Roto gen</i>	<i>Roto spec</i>	<i>Sto gen</i>	<i>Sto spec</i>	<i>Roto gen</i>	<i>Roto spec</i>	<i>Sto gen</i>	<i>Sto spec</i>
<b>landcover</b>												
<i>permanent pasture</i>	2%	3%	4%	NA	0%	NA	7%	NA	4%	15%	9%	13%
	11%				8%				1%			
<i>temporary pasture</i>	-12%	4%	-16%	5%	3%	8%	-2%	10%	18%	10%	13%	1%
	40%				23%				5%			
<i>silage maize</i>	-9%	-2%	-11%	-4%	0%	-4%	-2%	-2%	3%	1%	1%	-2%
	19%				13%				3%			
<i>grain maize</i>	11%	-2%	13%	-2%	2%	2%	2%	-1%	-13%	-11%	-11%	-1%
	5%				14%				36%			
<i>wheat</i>	6%	0%	7%	4%	-1%	2%	0%	0%	-6%	-10%	-6%	-8%
	17%				28%				38%			

Table III.3: Difference of crop proportions of the farm-type area between the 40-year simulations and the observed dataset (observed crop proportions are in italics). The difference was calculated with the values simulated by: Roto gen=generic matrix model, Roto spec=farm-type specific matrices model, Sto gen=generic stochastic decision trees, Sto spec=farm-type specific stochastic decision trees. Shaded cells indicate configurations where the farm-type expected a specific crop proportion (expected value are in italics). Bold numbers indicate that the 40-years series of simulated area proportions were significantly different (Wilcoxon test,  $p=0.95$ ) from the 1993-2002 series of observed area proportions.

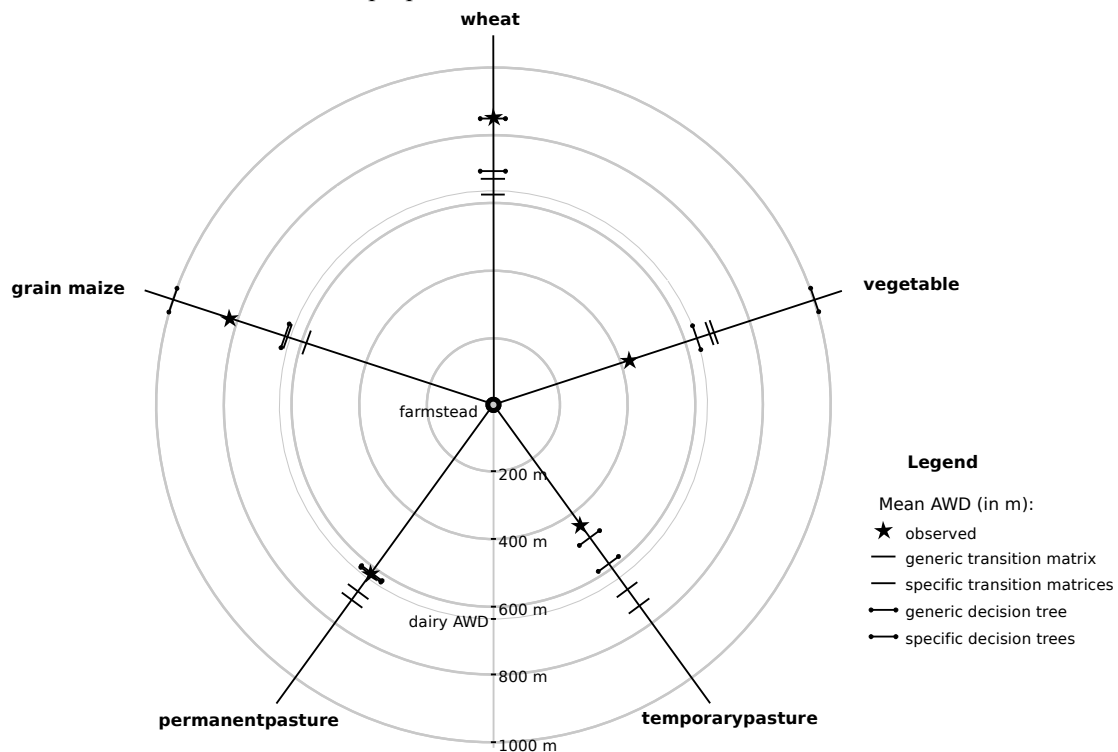


Fig. III.25: Modification of the spatial distribution of the crops around the farmstead for the dairy farm-type, for each crop transition models (CTM) measured on 40-year simulation (as if viewed from the sky). The spatial indicator is the area-weighted distance (AWD) of the crop plot pattern. Each crop is shown with its observed mean AWD (computed over 1993-2002) and the mean AWD simulated for 40 years by each CTM. All the differences between the simulated and observed AWD were significant (Wilcoxon test,  $p=0.95$ ) except for specific decision trees with wheat and grain maize.

### **Spatial distribution of the crops around the farmstead**

Fig. III.25 shows the modifications implied by the simulations to the crop spatial distributions for the dairy farm-type, which exhibited the most spatially-structured crop pattern in the observation dataset (Fig. III.21). The Wilcoxon comparison tests between the observed and the simulated series of AWD showed significant differences for all the simulations and on almost all the spatially structured crops (grain maize, wheat, permanent and temporary pastures, and vegetable). Significant differences between observed and simulated AWD ranged between 50 m and 300 m for the main crops of the dairy farm-type. *Stochastree* generally outperformed *Rotomatrix* given the same implementation (generic or specific). For a given crop, specific SDT caused smaller AWD differences than generic SDT. The figure also shows that *Rotomatrix* and, to a lesser extent *Stochastree*, tend homogenize crop AWD to the dairy farm-type AWD (636 m, Fig. III.21).

### **Preferential allocation of crops over the soil waterlogged class**

Each year, the area of each crop allocated on the waterlogged soil class was expressed as a percentage of the total crop area. Without consideration of the farm-type, these series of simulated proportions were compared to the observed series of proportions. Table III.4 shows the ability of the transition models to maintain the rate of occupation of a crop on the waterlogged soil class. For a given implementation of the model (generic or specific), the absolute differences with the observed proportions were slightly smaller for *Stochastree* than *Rotomatrix*. However, for a given type of model (*Rotomatrix* or *Stochastree*), the specific implementation showed generally larger significant differences than their generic implementation.

Beside permanent pastures, which stay on the same fields by definition, observed data on Naizin showed that waterlogged soils were mainly occupied by fallow, which exhibited the most striking differences of proportion between simulations in the favor of specific rather than generic implementations, and more generally, *Stochastree* over *Rotomatrix*.

crop class	simulation method			
	<i>Roto gen</i>	<i>Roto spec</i>	<i>Sto gen</i>	<i>Sto spec</i>
<i>permanent pasture</i>	<b>-6%</b>	<b>-6%</b>	<b>-8%</b>	<b>-6%</b>
		25%		
<i>temporary pasture</i>	-1%	<b>-3%</b>	-1%	<b>-4%</b>
		8%		
<i>fallow</i>	<b>-43%</b>	<b>-39%</b>	<b>-24%</b>	-9%
		52%		
<i>seed ray-grass</i>	NA	8%	-2%	-1%
		5%		
<i>silage maize</i>	2%	<b>2%</b>	1%	1%
		3%		
<i>grain maize</i>	-2%	-1%	-2%	-1%
		9%		
<i>wheat</i>	<b>3%</b>	<b>3%</b>	1%	<b>3%</b>
		4%		
<i>vegetable</i>	1%	1%	-2%	-1%
		5%		
<i>rapeseed</i>	7%	10%	6%	4%
		0.3%		
<i>potato</i>	2%	<b>6%</b>	1%	<b>6%</b>
		5%		

Table III.4: Difference of crop area proportion located on the waterlogged soil class between the 40-years simulation period and the 1993-2002 observation period (observed proportions are in italics). Bold numbers indicate that the difference between simulated and observed proportions is significant (Wilcoxon test,  $p=0.95$ ). The simulation methods are: Roto gen=generic Markov-chain model, Roto spec=farm-type specific Markov-chain model, Sto gen=generic stochastic decision trees, Sto spec=farm-type specific stochastic decision trees.

## Discussion

### ***Comparing observations and simulations***

The preliminary analysis of data collected on the Naizin catchment during 1993-2002 allowed us to quantify some features of the spatial allocation of management practices in farming landscapes. The dairy farm-type exhibited the tightest field pattern (AWD = 636 m), which results from accumulated constraints related to the field distance-to-farmstead: daily trips of dairy cows between the pastures and the milking room, water and food supplies to grazing heifers. In Naizin, pigs are bred in buildings associated with the farmstead, thus this farm-type require farmstead-field trips only for crop management and explains why the pig farm-type had a wider AWD (798 m). Although it has already been observed that (i) farm-types have specific crop objectives and (ii) crop management influences the crop pattern around the farmstead (Thenail and Baudry, 2004), the challenge of the decision trees construction was to expect that these driving factors would remain noticeable when mining the dataset at the field scale (the decision tree growth phase). Although Fig. III.23.a showed that all the related field attributes (soil waterlogging, field area, and field distance-to-farmstead) were actually incorporated in the decision trees, the different agronomic constraints were not fulfilled with the same level of satisfaction. Each driving factor is discussed hereafter.

### **Crop production objectives**

Actually, the root node of all the decision trees tested the “current crop” attribute, which made them very similar to Markovian transition matrices, whose mathematical properties (no negative term in the matrix, ergodicity) are known to induce or maintain stationarity of crop proportions when used to simulate crop transitions (Usher, 1992; Logofet and Lesnaya, 2000; Castellazzi et al., 2007b). The trained transition matrices had such properties and the observed stability of crop proportions during the 1993-2002 period (Fig. III.20) suggests that the system already was at a stationary state. The structural similarity of the stochastic decision trees, and the fact that more than half of decision leaves only involved test on the current crop, explain why *Stochastree* and *Rotomatrix* showed similar performances in terms of crop production objectives. Stratifying the learning dataset by farm-type allowed to build specific models, which simulated crop proportions closer to the observed ones. However the stationarity of the observed crop areas induced a high sensitivity in the Wilcoxon tests when comparing the observed and the simulated crop proportions: a difference of 3% (in absolute value) of the farm-type total area revealed to be statistically significant but may be acceptable from an agronomic perspective (Table III.3). However, the common structure of SDT and matrices made *Stochastree* simulations also sensitive to “absorbing” crops (Usher, 1979): the area dedicated to permanent pastures kept on increasing but at a very slow speed (around 0.5 ha.year<sup>-1</sup>) since the only transition leading to permanent pastures was from the vegetable crop class, which was very marginally represented in Naizin (Fig. III.20) and with low probabilities of apparition (Fig. III.22).

### **Spatial distribution of crops over the field pattern**

Fig. III.23.a showed that only few branches of the SDT involved the distance attribute in the process of simulating the field next crop, which makes the decision trees unlikely to maintain the observed spatial structures. This may explain the homogenization trend of the crop AWD towards the farm-type AWD observed in the simulations (Fig. III.25). However, Fig. III.23.a shows that the major crops (wheat, maize) were associated with the distance attribute and Fig. III.22 shows that they were also pointed by the highest transition probabilities, which suggests that distance may be solicited often, even though not for every crop transition. This explains why *Stochastree* modified the crop AWDs to a lesser extent than *Rotomatrix* (dairy farm-type, Fig. III.25).

### **Preferential allocation of crops over the soil waterlogged class**

The comparisons tests concerning the proportion of crop areas over the waterlogged soil class could hardly discriminate the LCM approaches, except for fallows, for which the use of specific SDT enabled to allocate them preferentially on waterlogged soils, as observed in reality. The fact that matrices were able to show the same performances than SDT on several crops is not so surprising for two reasons. First, soil waterlogging is a constant property of the landscape over the observation period (1993-2002). Therefore, it is very unlikely to observe transitions between resistant and sensitive crops, and this may have resulted in null or marginal transition probabilities in the matrices. Second, the drainage systems set up by the farmers may have caused local discrepancies in the learning dataset, with sensitive crops actually able to grow on soils mapped waterlogged.

### **Validation and generalization of *Stochastree* simulations**

Compared to the construction of crop transition models like ROTAT (Dogliotti et al., 2003) and ROTOR (Bachinger and Zander, 2007), our approach replaced expert knowledge by a data-mining phase to grow the decision trees, which required a substantial learning dataset. However, the construction of decision trees is known to be sensitive to the quality and the noisiness of learning dataset (Breiman et al., 1984), therefore a validation of the decision trees is necessary to assess their robustness. In this purpose, we compared the simulation results with an observed dataset with tests relying on non-parametric hypotheses, but implying time-invariance of the agronomic driving factors. This assumption of time-invariance of the driving factors is strong but realistic if the socio-economic factors like crop market values and agricultural policies do not change (Aspinall, 2004).

The use on a catchment of decision trees learned on a different one may also be biased if the construction of the DT was subject to overfitting of the learning dataset (Breiman et al., 1984; Murthy et al., 1998). Local configurations of fields, farmsteads and the road network may be a significant source of bias which can be avoided by adjusting the stop rule in the DT growing phase and parameterizing the pruning phase. However, concerning dairy cows, the amount of distance they can walk between pastures and the milking room may not differ a lot.

## ***Mutual benefits of decision trees and Markov chains approaches in LCM***

The Naizin catchment is mainly occupied by breeding farms (dairy cows or pigs), therefore a large proportion of the cultivated area is dedicated to forage and cereals to provide fodder and pastures. The analysis of the 1993-2002 landcover data confirmed the key role of these crops by showing high production objectives of forage for dairy farms and grain maize for pig farms. Conversely, only small areas were generally dedicated to market crops (vegetable and potato) and no farm-type exhibited significant production objective for them. The temporal Markovian approach illustrated graphically the importance of cereals in the crop transitions and also helped to analyze the relative importance of test nodes and the conditional probabilities in decision trees. The box and arrow representation of crop transitions also highlighted a potential issue of crop transition models built with learning methods: the existence of “absorbing” crops.

To account for particular configurations and thus enabling more flexibility, several LCM based on Markov chains stratify the transition probabilities on landscape features like slope, waterlogging or the current crop at the vicinity of the area of interest (Weaver and Perera, 2003; Ladet et al., 2005; Pocewicz et al., 2008). The use of decision trees could preliminary suggest the attributes on which it is relevant to stratify these probabilities by looking at the attributes kept in the tree structure.

### ***Improving the construction of the decision trees***

Although decision trees are mainly used as classifier in the field of landcover analysis and simulation, issues concerning the estimation of the class probabilities at the DT leaves started very early (Breiman et al., 1984) and are still debated (Provost and Domingos, 2003; Liang et al., 2006). One of the recurrent critics to algorithms that construct iteratively decision trees concern the “divide and conquer” method (Murthy et al., 1998): by splitting iteratively the learning dataset into smaller partitions, the decisions are based on increasingly smaller fragments of the data. In this paper, we built the farm-type specific DT by stratifying the learning dataset by farm-type and generating each DT individually. This first fragmentation division of the dataset followed by the “current crop” test of the dataset may be the reasons why the dairy-pig has only one child leaf testing other attributes than the “current crop” (Fig. III.23.a). To avoid the progressive loss of learning instances, some researchers introduced the use of soft splits allowing some instances to propagate in the child branches with associated probabilities (Heidenberger, 1996; Jordan et al., 1997; Hazen, 2002) inspired by the fuzzy logic theory (Yuan and Shaw, 1995). Once such a tree is built, soft splits offer flexibility to handle (i) instances of noisy dataset and (ii) sparse instances. For example, such fuzzy DT would probably handle in a better way instances characterized by a hydrophobic crop associated with a waterlogged soil class, by assuming that there must have been an error either in labeling the crop or the waterlogging class. In the same way, recent papers have pointed out the need for fuzzy hierarchical classifications of landcover to handle ambiguous information in remotely sensed data (Rocchini and Ricotta, 2007), obviously, LCM must incorporate some fuzziness processes to be coupled to such approaches.



The decision trees used in this research belong to the family of the so-called univariate DT, since a test node only test one attribute of the instance. There is also another family of multi-variate DT (Friedl and Brodley, 1997; Pal and Mather, 2003) where multiple descriptors can be combined to stratify the dataset (combining landcover and waterlogging for example, or area and distance, etc.) in new efficient ways, although multi-variate approaches still require parameter tuning and expert calibration (Friedl and Brodley, 1997).

## Conclusion

In this paper, we described and assessed the relevance of an innovative landcover change model based on stochastic decision trees, which proved to be relatively efficient in maintaining an observed spatial structure of the crops, while reasonably following the main production objectives and the preferential allocation of grasslands on waterlogged soil. The ability to consider such driving factors and to maintain such spatial structures in the simulation outputs is a significant added value for the coupling of *Stochastree* with nutrient or pesticide fate models. *Stochastree* could also be coupled with other types of model requiring spatially distributed landcover. It could simulate entries for models of gene contamination, for which the temporal and spatial allocation of the crops in the fields influences the simulation outputs (Castellazzi et al., 2007a). However, the relevance of using *Stochastree* in such a case would involve other tests concerning landcover connectivity between fields.

This first application opens doors to potential new uses of decision tree methods in the field of landcover modeling. Most of the landcover change models require an initial landcover state to simulate the landcover successions, but stochastic decision trees could also be trained by removing the “current landcover” attribute in the learning dataset and used to model the initial state of a landscape where landcover is unknown or incomplete. This type of use is emerging in the supervised processing of remotely sensed data (Pal and Mather, 2003), and would offer a pragmatic or empirical alternative to neutral landscapes model (Gaucherel et al., 2006b) which are built without considering specific ecological processes (Turner et al., 2001).

## Acknowledgments

The PhD of Luc Sorel was funded by INRA and Région Bretagne. We are grateful to Paul Bordenave (Cemagref) who provided the landcover data of Naizin, to Jordy Salmon-Monviola and Ricco Rakotomalala (Université de Lyon 2) for stimulating discussions.

## References

- Agarwal C., G.M. Green, J.M. Grove, T.P. Evans, C.M. Schweik. 2002. A review and assessment of land-use change models: dynamics of space, time, and human choice. Gen. Tech. Rep. NE-297. Newton Square, PA: U.S. Department of Agriculture, Forest Service, Northeastern Research Station. 61 p.
- Anderson T.W. and L.A. Goodman. 1957. Statistical inference about Markov chains. *Annals of Mathematical Statistics*, 28 (1):89-110.
- Aspinall R. 2004. Modelling land use change with generalized linear models – a multi-model analysis of change between 1860 and 2000 in Gallatin Valley, Montana. *Journal of Environmental Management*, 72:91-103.
- Bachinger J. and P. Zander. 2007. ROTOR, a tool for generating and evaluating crop rotations for organic farming systems. *European Journal of Agronomy*, 26 (2):130-143.
- Baker W.L. 1989. A review of models of landscape change. *Landscape ecology*, 2 (2):111-133.
- Baudry J., C. Dupont, C. Thenail, V. Viaud. 2006. Aménagement du paysage et pratiques agricoles : quelles combinaisons dans la gestion des bassins versants ? *In Mérot, P. Qualité de l'eau en milieu rural : Savoirs et pratiques dans les bassins versants*. Inra Eds, Paris:123-135.
- Beaujouan V., P. Durand, L. Ruiz. 2001. Modelling the effect of the spatial distribution of agricultural practices on nitrogen fluxes in rural catchments. *Ecological Modelling*, 137:93-105.
- Beaujouan V., P. Durand, L. Ruiz, P. Arousseau, G. Cotteret. 2002. A hydrological model dedicated to topography-based simulation of nitrogen transfer and transformation: rationale and application to the geomorphology-denitrification relationship. *Hydrological Processes*, 16:493-507.
- Beven K.J. 2001. *Rainfall-Runoff Modelling – The Primer*, Wiley, Chichester.
- Bockstael N., R. Costanza, I. Strand, W. Boynton, K. Bell, L. Wainger. 1995. Ecological economic modeling and valuation of ecosystems. *Ecological Economics*, 14:143-59.
- Bordenave P., F. Oehler, T. Bioteau, N. Turpin, P. Serrand, P. Saint-Cast, E. Le Saos. 2005. Distributed surface-groundwater coupled model applied to long term water management impacts at basin scale. *Ingénieries*, 44:13-36.
- Breiman L., J.H. Friedman, R.A. Olshen, C.J. Stone. 1984. *Classification and regression trees*. Monterey, CA, Wadsworth.
- Carpenter S.R., N.F. Caraco, D.L. Correll, R.W. Howarth, A.N. Sharpley, V.H. Smith. 1998. Nonpoint pollution of surface waters with phosphorus and nitrogen. *Ecological Applications*, 8 (3):559-568.
- Castellazzi M.S., J.N. Perry, N. Colbach, H. Monod, K. Adamczyk, V. Viaud, K.F. Conrad. 2007a. New measures and tests of temporal and spatial pattern of crops in agricultural landscapes. *Agriculture, Ecosystems & Environment*, 118 (1-4):339-349.
- Castellazzi M.S., G.A. Wood, P.J. Burgess, J. Morris, K.F. Conrad, J.N. Perry. 2007b. A systematic representation of crop rotations. *Agricultural Systems*, 97 (1-2):16-33.
- Cheverry C. (ed.). 1998. *Agriculture intensive et qualité des eaux*. Science Update, Paris. 297 pages.
- Coppedge B.O., D.M. Engle, S.D. Fuhlendorf. 2007. Markov models of land cover dynamics in a southern Great Plains grassland region. *Landscape Ecology*, 22:1383-1393.
- D'Arcy B. and A. Frost. 2001. The role of best management practices in alleviating water quality problems associated with diffuse pollution. *The Science of the Total Environment*, 265:359-367.
- DeFries R.S., J.C.W. Chan. 2000. Multiple criteria for evaluating machine learning algorithms for land cover classification from satellite data. *Remote Sensing of Environment*, 74 (3):503-515.
- Dogliotti S., W.A.H. Rossing, M. K. van Ittersum. 2003. ROTAT, a tool for systematically generating crop rotations. *European Journal of Agronomy*, 19 (2):239-250.
- Friedl M.A. and Brodley, C.E. 1997. Decision tree classification of land cover from remotely sensed data. *Remote Sensing of Environment*, 61 (3):399-409.
- Friedl M.A., D.K. McIver, J.C.F. Hodges, X.Y. Zhang, D. Muchoney, A.H. Strahler, C.E. Woodcock, S. Gopal, A. Schneider, A. Cooper, A. Baccini, F. Gao and C. Schaaf. 2002. Global land cover mapping from MODIS: algorithms and early results. *Remote Sensing of Environment*, 83 (1-2):287-302.

- Gabrielle B., B. Mary, R. Roche, P. Smith, G. Gosse. 2002. Simulation of carbon and nitrogen dynamics in arable soils: a comparison of approaches. *European Journal of Agronomy*, 18:107-120.
- Gaucherel C., N. Giboire, V. Viaud, T. Houet, J. Baudry, F. Burel. 2006a. A domain-specific language for patchy landscape modelling: the Brittany agricultural mosaic as a case study. *Ecological Modelling*, 194:233-243.
- Gaucherel C., D. Fleury, D. Auclair, P. Dreyfus. 2006b. Neutral models for patchy landscapes. *Ecological Modelling*, 197 (1-2):159-170.
- Gladwin C.H. 1989. *Ethnographic decision tree modeling*. Sage, Newbury Park.
- Grizzetti B., F. Bouraoui, G. De Marsily. 2005. Modelling nitrogen pressure in river basins: a comparison between a statistical approach and the physically-based SWAT model. *Physics and Chemistry of the Earth*, 30 (8-10):508-517.
- Hanink D.M. and R.G. Cromley. 1998. Land-use allocation in the absence of complete market values. *Journal of Regional Science*, 38 (3):465-480.
- Hazen G.B. 2002. Stochastic tree and the StoTree modeling environment: models and software for medical decision analysis. *Journal of Medical Systems*, 26 (5):399-413.
- Heidenberger K. 1996. Dynamic project selection and funding under risk: a decision tree based MILP approach. *European Journal of Operational Research*, 95:284-298.
- Houet T. and L. Hubert-Moy, 2006. Modelling and projecting land-use and land-cover changes with cellular automaton in considering landscape trajectories: an improvement for simulation of plausible future states. *EARSeL eProceedings*, 5 (1):63-76.
- Huang Q.H., Y.L. Cai, J. Peng. 2007. Modeling the spatial pattern of farmland using GIS and multiple logistic regression: a case study of Maotiao River Basin, Guizhou Province, China. *Environmental Modeling and Assessment*, 12 (1):55-61.
- Jordan M.I., Z. Ghahramani, L.K. Saul. 1997. Hidden Markov decision trees. *In* M. C. Mozer, M.I. Jordan, T. Petsche (Eds.), *Advances in Neural Information Processing Systems (NIPS) 10*, Cambridge MA: MIT Press.
- Karpinets T.V., D.J. Greenwood, J.T. Ammons. 2004. Predictive mechanistic model of soil phosphorus dynamics with readily available inputs. *Soil Science Society of America Journal*, 68 (2):644-653.
- Ladet S., M. Deconchat, C. Monteil, J.P. Lacombe, G. Balent. 2005. Les chaînes de Markov spatialisées comme outil de simulation. *Géomatique, Cassini*, 04:159-173.
- Largouët C. and M.O. Cordier. 2001 Improving the landcover classification using domain knowledge. *AI Communication*, 1:35-43.
- Lawrence R., A. Bunn, S. Powell, M. Zambon. 2004. Classification of remotely sensed imagery using stochastic gradient boosting as a refinement of classification tree analysis. *Remote Sensing of Environment*, 90:331-336.
- Le Ber F., M. Benoit, C. Schott, J.F. Mari, C. Mignolet. 2006. Studying crop sequences with CarrotAge, a HMM-based data mining software. *Ecological Modelling*, 191 (1):170-185.
- Lewis D.R. and M.B. McGehan. 2002. A review of field scale phosphorus dynamic models. *Biosystems Engineering*, 82 (4):359-380.
- Li X. and A.G.O. Yeh. 2002. Neural-network-based cellular automata for simulating multiple land use changes using GIS. *International Journal of Geographical Information Science*, 16 (4):323-343.
- Liang H., H. Zhang, Y. Yan. 2006. Decision Trees for Probability Estimation: An Empirical Study. NRC 48783. The 18th IEEE International Conference on Tools with Artificial Intelligence (ICTAI06). Washington D.C., USA. November 13-15, 2006. 10 pages.
- Logofet D.O. and E.V. Lesnaya. 2000. The mathematics of Markov models: what Markov chains can really predict in forest successions? *Ecological Modelling*, 126:285-298.
- Lynn H., C.L. Mohler, S.D. DeGloria, C.E. McCulloch. 1995. Error assessment in decision-tree models applied to vegetation analysis. *Landscape Ecology*, 10 (6):323-335.
- Mignolet C., C. Schott, M. Benoît. 2004. Spatial dynamics of agricultural practices on a basin territory: a retrospective study to implement models simulating nitrate flow. The case of the Seine basin. *Agronomy for Sustainable Development*, 24:219-236.
- Molénat J. and C. Gascuel-Oudou. 2002. Modelling flow and nitrate transport in groundwater for the prediction of water travel times and of consequences of land use evolution on water quality. *Hydrological Processes*, 16 (2):479-

- Murthy S.K. 1998. Automatic construction of decision trees from data: a multi-disciplinary survey. *Data Mining and Knowledge Discovery*, 2:345-389.
- Pal M. and P.M. Mather. 2003. An assessment of the effectiveness of decision tree methods for land cover classification. *Remote Sensing of Environment*, 86:554-565.
- Pauwels H., P. Lachassagne, P. Bordenave, J.C. Foucher, A. Martelat. 2001. Temporal variability of nitrate concentration in a schist aquifer and transfer to surface waters. *Applied Geochemistry*, 16:583-596.
- Payraudeau S., H.M.G. van Der Werf, F. Vertès. 2006. Evaluation of an operational method for the estimation of emissions of nitrogen compounds for a group of farms. *International Journal of Agricultural Resources, Governance and Ecology*, 5 (2-3):224-246.
- Perlich C., F. Provost, J. S. Simonoff. 2003. Tree induction vs. logistic regression: a learning-curve analysis. *Journal of Machine Learning Research*, 4 (2):211-255.
- Pocewicz A., M. Nielsen-Pincus, C. Goldberg, M. Johnson, P. Morgan, J. Force, L. Waits, L. Vierling. 2008. Predicting land use change: comparison of models based on landowner surveys and historical land cover trends. *Landscape Ecology*, 23 (2):195-210.
- Provost F. and P. Domingos. 2003. Tree induction for probability-based ranking. *Machine Learning*, 52 (3):199-215.
- Quinlan J. R. 1981, Induction of decision trees. *Machine Learning*, 1:81-106.
- Quinlan J.R. 1990. Probabilistic decision trees. *In Machine Learning: An Artificial Intelligence Approach - Volume III*, Y. Kodratoff and R.S. Michalski, Eds. Morgan Kaufmann, San Mateo, CA, 140-152.
- Quinlan J.R. 1993. *C4.5: Programs for Machine Learning*. San Francisco: Morgan Kaufmann.
- Rocchini D. and C. Ricotta. 2007. Are landscapes as crisp as we may think? *Ecological Modelling*, 204:535-539.
- Rounsevell M.D.A., J.E. Annetts, E. Audsley, T. Mayr, I. Reginster. 2003. Modelling the spatial distribution of agricultural land use at the regional scale. *Agriculture, Ecosystems & Environment*, 95 (2-3):465-479.
- Russel E.W. 1973. *Soil condition and plant growth*. 10<sup>th</sup> ed. Longmans, London.
- Shannon C.E. 1948. A mathematical theory of communication. *Bell System Technical Journal*, 27:379-423, 623-656.
- Tarim S.A., S. Manandhar, T. Walsh. 2006. Stochastic constraint programming: a scenario-based approach. *Constraints*, 11 (1):53-80.
- Thornton P.K. and P.G. Jones. 1998. A conceptual approach to dynamic agricultural land-use modelling. *Agricultural Systems*, 7:505-521.
- Thenail C. and J. Baudry. 1994. Méthodes d'étude des relations entre activités agricoles et paysages. *In Sébillotte M. (Ed.), Recherches-Systèmes en Agriculture et Développement Rural; 13e Symposium International*. CIRAD Editions, Montpellier:316-321.
- Thenail C. and J. Baudry. 2004. Variation of farm spatial land use pattern according to the structure of the hedgerow network (bocage) landscape: a case study in northeast Brittany. *Agriculture, Ecosystems & Environment*, 101:53-72.
- Turner M.G., R.H. Gardner, R.V. O'Neil. 2001. *Landscape ecology in theory and practice: Pattern and process*. Springer-Verlag, New York, USA.
- Usher M.B. 1979. Markovian approaches to ecological succession. *Journal of Animal Ecology*, 48:413-426.
- Usher M.B. 1992. Statistical models of succession. *In Glenn-Lewin DC, Peet RK, Veblen TT (eds). Plant succession: theory and prediction*. Chapman & Hall, London:215-248.
- Vadas P.A., R.D. Harmel, P. J. A. Kleinman. 2007. Transformations of soil and manure phosphorus after surface application of manure to field plots. *Nutrient Cycling in Agroecosystems*, 77:83-99.
- Veldkamp A. and L.O. Fresco. 1996. CLUE: a conceptual model to study the Conversion of Land Use and its Effects. *Ecological Modelling*, 85:253-270.
- Verburg P.H. 2006. Simulating feedbacks in land use and land cover change models. *Landscape Ecology*, 21:1171-1183.
- Verburg P.H., P.P. Schot, M. J. Dijst, A. Veldkamp. 2004. Land use change modelling: current practice and research priorities. *Geojournal*, 61:309-324.
- Walter C. and P. Curmi. 1998. Les sols du bassin versant du Coët-Dan : organisation, variabilité spatiale et cartographie.

- In* Cheverry C. (ed.). Agriculture intensive et qualité des eaux. Science Update: Paris:85-105.
- Walter C., R.A. Visccarra Rossel, A.B. McBratney. 2003. Spatio-temporal simulation of the field-scale evolution of organic carbon over the landscape. *Soil Science Society of America Journal*, 67:1477-1486.
- Weaver K. and A.H. Perera. 2003. Modelling land cover transitions: A solution to the problem of spatial dependence in data. *Landscape Ecology*, 19 (3):273-289.
- Witten I.H. and E. Frank. 2005. *Data Mining: Practical machine learning tools and techniques*, 2<sup>nd</sup> Edition. Morgan Kaufmann, San Francisco.
- Wu F. and C.J. Webster. 1998. Simulation of land development through the integration of cellular automata and multicriteria evaluation. *Environment and Planning B: Planning and Design*, 25 (1):103-126.
- Yuan Y. and M.J. Shaw. 1995. Induction of fuzzy decision trees. *Fuzzy Sets and Systems*, 69 (2):125-139.
- Zhang C. and W. Li. 2008. A comparative study of nonlinear Markov chain models for conditional simulation of multinominal classes from regular classes. *Stochastic Environmental Research and Risk Assessment*, 22 (2):217-230.

## **Chapitre IV**

Une approche multi-ressources et multi-critères à l'échelle  
du paysage, d'évaluation de l'impact environnemental des  
pratiques agricoles



# Chapitre IV : Une approche multi-ressources et multi-critères à l'échelle du paysage, d'évaluation de l'impact environnemental des pratiques agricoles

## Résumé français

Les démarches d'évaluation de l'impact des pratiques agricoles sur les ressources deviennent progressivement plus intégratives, ne se focalisant plus sur un seul polluant ou une seule ressource : seules des approches, considérant de façon simultanée les différentes ressources et les différentes voies de dégradation des milieux, apparaissent en effet à même d'identifier et de limiter les éventuels transferts de polluants et d'assurer la mise au point de systèmes de production agricole durables (Payraudeau et Van der Werf, 2005) ou de bonnes pratiques agricoles (Carpenter *et al.*, 1998 ; Hart *et al.*, 2004).

L'objectif général de ce chapitre est de développer une démarche de modélisation dynamique et spatialisée à l'échelle du paysage, qualifiée de paysage virtuel, pour évaluer l'impact environnemental de scénarios agronomiques selon les conditions du milieu. Les ressources considérées sont les ressources en eau (nappes, cours d'eau) et en sols et l'accent est mis sur les flux d'azote et de phosphore.

Trois objectifs spécifiques sont identifiés : (i) concevoir un ensemble multifactoriel de paysages virtuels issus de la modification de la structure spatiale des sols, du climat et pratiques agricoles observés sur un paysage réel ; (ii) coupler ces paysages virtuels au modèle TNT2 de transformation et de transfert de l'azote (Beaujouan *et al.*, 2002) et à un modèle spécifique de bilan de phosphore dans les sols ; (iii) estimer les effets relatifs et les interactions des composantes liées aux sols, au climat et aux systèmes de culture, sur l'évolution des ressources.

La démarche repose sur la construction de 18 paysages virtuels dérivés de la caractérisation détaillée du bassin versant de Kervidy (Naizin, Morbihan) d'une superficie de 500 ha, dont les systèmes de production agricole sont principalement orientés vers l'élevage intensif de porcs ou la production laitière. Ces paysages virtuels résultent de la combinaison de deux systèmes de culture (intensif, modéré), de trois séries climatiques et de trois configurations de l'épaisseur et de l'hydromorphie des sols. Un paysage virtuel de référence est celui correspondant au système de culture intensif actuellement adopté sur le bassin versant et aux conditions pédologiques et climatiques observées.

La comparaison démarre d'un même état initial, issu d'une simulation préalable sur une durée de six ans de TNT2 sur le paysage virtuel de référence, permettant ainsi d'initialiser la distribution des paramètres des modèles. Chacun de ces 18 paysages virtuels sert ensuite successivement de jeu de données d'entrée du modèle TNT2 et du modèle de bilan de phosphore, sur une période de simulation de 40 ans au cours de laquelle l'évolution de l'occupation du sol est simulée par le modèle *Stochastree* (chapitre III).



La comparaison des scénarios est fondée sur des statistiques annuelles ou des cumuls estimés à partir des 15 dernières années de simulation, en faisant l'hypothèse que les bassins versants ont atteint un état d'équilibre au bout de 25 ans. Vingt indicateurs agronomiques, pédologiques et hydrologiques ont été retenus et analysés par analyse en composantes principales (ACP) et analyse de variance pour évaluer les effets des systèmes de culture, du climat, des sols et de leurs interactions.

Les résultats montrent que les indicateurs retenus sont en premier lieu influencés par le climat, puis par les systèmes de culture, enfin par les sols. L'adoption d'un système de culture modéré réduit significativement les flux excédentaires de phosphore vers les sols et accroît l'efficacité des apports de N et P ; néanmoins les configurations de climat et de sol influencent davantage les concentrations en nitrate des cours d'eau que les pratiques. De plus, les effets des interactions sol-climat sur cette concentration sont du même ordre qu'une réduction de 15 % des apports totaux d'azote. Cela montre bien que la définition de bonnes pratiques agricoles ne peut se faire qu'en lien avec les conditions du milieu et que leur transposition d'un bassin versant à l'autre peut s'accompagner de variations importantes de leur efficacité.

Cette étude propose ainsi une démarche innovante d'évaluation multi-ressources et multi-critères des impacts environnementaux des pratiques agricoles, dont l'extension est envisagée pour intégrer d'autres ressources (eau, énergie) et d'autres critères (matière organique, gaz à effet de serre, etc.).

# Multi-resource and multi-criteria landscape approach in assessing the environmental impacts of agricultural practices.

Luc Sorela,b, Patrick Durand\*a,b, Valérie Viauda,b, Christian Waltera,b

<sup>a</sup> INRA, UMR1069, Sol Agro et hydrosystème Spatialisation, F-35000 Rennes

<sup>b</sup> Agrocampus Rennes, UMR1069, Sol Agro et hydrosystème Spatialisation, F-35000 Rennes

\* Corresponding author: patrick.durand@rennes.inra.fr INRA-Agrocampus Rennes, UMR SAS, 65 rue de Saint Briec, CS 84215, 35042 Rennes Cedex, France (phone: 0(+33) 2-23-48-59-42 / fax: 0(+33) 2-23-48-54-30)

Abbreviations: AWC, available water content; DEM, digital elevation model; DM, dry matter; DTI, downstream topographic index; EU, European Union; NST, normal score transform; PET, potential evapotranspiration; N, nitrogen; P, phosphorus.

## Abstract

Mitigation of non-point agricultural pollution related to nutrient is generally based on to the recommendation of best management practices (BMP) concerning crop, fertilization and waste management. However, climate and soil govern nutrient transformation and transfer, and the related variability has to be quantified before implementing BMP over a wide diversity of farming catchments. In order to assess the relative influence of climate, physical environment, and cropping systems on soil and water quality, we designed a multi-factorial set of virtual landscapes derived from a real monitored farming catchment. Eighteen virtual landscapes were designed as combination of two cropping systems, three soil climates series, and three soil pattern configurations. The landscapes served as entries to a nitrogen (N) distributed fate model and a soil phosphorus (P) budget model. Agricultural, soil and hydrological indicators were analyzed through PCA and ANOVA methods to evaluate the effects of practice, climate, soil factors and pairwise interactions. Environmental indicators were driven firstly by climate, practice, then soil levels. Although the moderate cropping system reduced satisfactorily soil P surpluses and improved N and P use efficiencies, climate and soil pattern influenced stream N concentration more heavily than cropping system. Moreover, the magnitude of soil-climate interactions on stream N concentrations was similar to the outcome of a 15% reduction of the total N inputs. This suggests that BMP designed on a specific area may be over- or under-efficient when transposed to other catchments.

## Introduction

In order to provide food in larger quantities or at lower prices, intensive agricultural practices have been developed and submitted soils to high crop yield expectations and heavy loads of nutrient inputs. Relations have been established between such practices and environmental impacts such as eutrophication of freshwaters mainly due to nitrogen (N) and phosphorus (P) runoff or leaching (Carpenter et al., 1998; Sharpley et al., 2000; Hart et al., 2004), soil organic carbon depletion with consequent erodibility (Lal, 2004), emission of greenhouse gases (GHG) like carbon dioxide and nitrous oxide (Hutchings et al., 2007). Existing regulations to preserve environmental resources, like the European Union (EU) water framework directive (European Commission, 1991), use two kinds of incentive to foster more sustainable practices: considering the freshwater resource itself, its N concentration must be lower than  $11.3 \text{ mg N.l}^{-1}$ ; considering cropping systems, N inputs must be balanced with crop needs. Regional regulations in vulnerable zones may also limit the annual P inputs to agricultural soils, e.g. an upper limit of  $44 \text{ kg P.ha}^{-1}$  was set in Ille-et-Vilaine (Brittany, France) (Préfecture d'Ille et Vilaine, 2005), and of  $9 \text{ kg P.ha}^{-1}$  in the Netherlands (Aarts et al., 2000). A wide range of practice assessment surveys conducted at various time and spatial scales provided insights on the mechanisms of nutrient fate (Heathwaite et al., 2003; Basset-Mens et al., 2006; Karlen et al., 2006; Oehler et al., 2007; Mignolet et al., 2007), which can be used to design BMP. But although they provide comprehensive frameworks to address remediation assessment issues, most studies conclude that the calibration of the methods is site-specific, because of the heterogeneity of landscape geomorphology and the diversity of its soil, climate, and agricultural components. This raises a critical issue in the design of best management practice (BMP): before being implemented region- or nation-wide, how can we ensure that they are adapted to the geomorphological and climate contexts where environmental resources have to be protected?

Nutrient transfers to stream water involve horizontal and vertical fluxes whose extent encompasses the whole catchment. Developing an experimental approach at this scale to detect and monitor the main processes appears infeasible for three main kinds of reasons. Firstly, the hydrological time response of catchments can exceed several years (Pauwels et al., 2001; Molénat et al., 2002) while the dynamics of nutrient transformation can be slow (Gabrielle et al., 2002), particularly the residual effects of manure practices on soil P, which can persist over decades (Russel, 1973; Karpinets et al., 2004). Secondly, specific factorial combinations involving climate or human activities cannot be neither controlled nor implemented on demand (Gustafson and Crow, 1998). Thirdly, the uniqueness of a catchment and the unknowability of its structure (Beven, 2002) complicate site comparisons or hypothesis verification (Ruiz et al., 2006). As a consequence, wide range of possible explanations can be advocated to explain that discharges or load dynamics expected by modeling hypotheses are not observed on field: effect of hedgerows, ditches, bedrock faults, disturbed underground water pathways, etc. Virtual landscapes, which are modeled and simplified representations of landscape, and spatial modeling have been used as remedy to this epistemological issue in several disciplines such as: hydrology (Beaujouan et al., 2001; Beaujouan et al., 2002), soil science (Walter et al., 2003),

ecology (Ruiz et al., 2006) and forestry management (Gustafson and Crow, 1998). The use of virtual landscapes allows to test hypotheses by deliberately setting the variability of chosen landscape characteristics, provided with the restriction that results interpretation should match the simplifications done and the explanatory depth of the approach (Beven, 2002).

Virtual landscapes can be coupled to dynamic and distributed models to simulate their spatio-temporal evolution. Although several N fate models exist, the topography-based nitrogen transfer and transformation model (TNT2) was specially designed to study the catchment response to distributed agricultural practices, as it simulates the possibility for the N leached upslope to be either stored in the vadose zone, assimilated downslope by plant uptake, or denitrified in valley bottom wetlands after transfer through the groundwater (Beaujouan et al., 2001; Beaujouan et al., 2002). Several field-scale P transformation and loss models have been developed but the heavy experiments required to calibrate them locally forbid them to be implemented at the catchment scale (Lewis and McGehan, 2002). Vadas et al. (2007) proposed recently a promising and simpler field-scale model designed to account for the influence of recent fertilizer applications on P losses with runoff. However, its implementation at the catchment level implies scale-change issues (P runoffs routing, deposition, and re-mobilization) which remain challenging issues (Sonneweld et al., 2006, Deasy et al., 2008). Moreover, annual P losses with water hardly exceeds  $2 \text{ kg P}\cdot\text{ha}^{-1}$  while input loads often exceed  $30 \text{ kg P}\cdot\text{ha}^{-1}$  (Brazier et al., 2005). Although very small concentrations of P can cause eutrophication in a waterbody, the amounts of P lost per annum with water fluxes are generally inconsequential from an agronomic perspective (Hart et al., 2004), thus the P pressure on soils can be estimated with a simple budget between P inputs and outputs by harvested crops and grazed or mowed grasslands.

In this paper, our objectives were (i) to design a multi-factorial set of virtual landscapes deriving from an extensively monitored small farming catchment; (ii) to couple them with TNT2 and a soil P budget model designed on purpose; (iii) to estimate the relative effects of cropping systems, climate, and soil patterns on soil and water quality. These three factors were chosen to design the virtual landscapes because soils are the support where crops grow, climate governs the hydrological fluxes, and cropping systems drives the amount of fertilizers spread on crops. The multi-criteria methodology proposed here involves agronomic and hydrological indicators such as N and P use efficiencies, comparison of N concentration with the  $11.3 \text{ mg N}\cdot\text{l}^{-1}$  EU permissible threshold, and input N transmission rate to the stream. Firstly, we present the main characteristics, hypotheses, data inputs and outputs of the models used to fulfill our objective. Then, we describe the methodology used to construct the virtual landscapes and to analyze the simulations.

# Materials and methods

## *Hydrological and nutrient fate models*

### **TNT2**

TNT2 is a process-based and integrative model (Beaujouan et al., 2002). Crop growth, water content, and N exchanges between crops and soils are simulated with the formalisms of STICS (Brisson et al., 1998). The denitrification module is an adaptation of NEMIS (Hénault and Germon, 2000) with two modifications: (i) the original water effect function was replaced by a function where a short (<5 days) water residence time (ratio between total drainage porosity volume and the new water passing through these soil layers each day) may partially inhibits denitrification; (ii) denitrification processes only occur in the top 20 cm layer. The hydrological module is an adaptation of TOPMODEL (Beven and Kirkby, 1979) and the four main assumptions are: (i) the hydraulic gradient is constant and equals the surface slope gradient, (ii) substratum hydraulic conductivity decreases exponentially with depth, (iii) saturated fluxes are described by Darcy's law, and (iv) overland flow is generated on saturated areas. The daily climate data used are rainfall and estimated potential evapotranspiration (PET) for the hydrological module, air temperature (minimum and maximum) and global radiation for the nitrogen transformation modules.

The model is fully distributed and relies on a regular square grid description of the landscape for topography, soil characteristics and field contours. The management of a crop is formalized in TNT2 by a set of cultural operations (tillage, sowing, mineral and organic N applications, harvesting) grouped in a crop management plan (CMP). For a given crop, several CMP may be used to account for different cropping systems, particularly concerning the amount and the nature of N applications. Each field has an associated management file which lists the succession of CMPs occurring on the field. For a given crop grown on a field of a given farm-type, the choice of the CMP used can simulate the choice of cropping systems (intensive, moderate, etc.). TNT2 is suited for relatively small catchments (1-30 km<sup>2</sup>) since all the fields must be individually characterized, and where hydrology is governed by shallow impermeable bedrock.

TNT2 outputs many variables, some at the catchment scale and at a daily time step (stream discharge, proportion of the catchment with soils saturated with groundwater, proportion of the catchment area with soils penetrated by the water table, stream and groundwater N concentrations and loads, mineralized and denitrified amounts), others at the field scale (harvested N and dry matter for crops, mowed and grazed dry matter and N for grasslands, and N excreta inputs to grazed grasslands).

### **Soil P budget model**

The soil P budget model designed for this study computes P inputs and outputs at the field scale, on the basis of N inputs, N outputs and dry matter (DM) inputs simulated by TNT2. Organic P inputs comprise manure spread on crops and excreta on grazed grasslands. The N and P composition for a wide range of manures and slurries from cattle, pig and poultry have been extensively studied by agricultural technical institutes, as well

as cattle excreta (Bodet et al., 2001), which gives the possibility to approximate P inputs by multiplying the N inputs by specific P/N ratios. Additional mineral inputs are essentially spread with starter fertilizers on vegetables, potatoes and maize after the sowing, and the estimation of applied amounts requires survey work. Concerning P outputs, nutrient contents have also been extensively monitored and the P/N ratio of mature crops at harvest proved to be relatively steady (Schvartz et al., 2005). Such ratios were therefore used to approximate P exports from the N exports estimated by TNT2. For grasslands, Schvartz et al. (2005) showed that P/N ratios change more widely depending on the level of maturity at which it is grazed and on the rate of fertilization. But the P/DM ratio was rather steady in time, thus mowed and grazed P amounts can be estimated from the DM inputs. The values used in this paper are presented in Table IV.5.

effluent	P/N
cattle excreta	0.07
cattle manure	0.18
cattle slurry	0.18
pig manure	0.47
pig slurry	0.30
poultry manure	0.42
poultry slurry	0.49

(a)

crop	associated N fertilization	mineral P amount (kg P,ha <sup>-1</sup> )
spinach	any	20
vegetable	any	20
maize	any organic	20
maize	mineral only	30
potato	any	60

(b)

harvested crop	P/N
wheat	0.19
maize	0.17
vegetable	0.04
rapeseed	0.18
potato	0.18
seed ray-grass	0.15
Italian ray-grass	0*
mustard	0*
grassland	0*

(c)

grassland inputs	P/DM
graze	2.13
mow	3.17

(d)

Table IV.5: Values used by the soil phosphorus (P) budget for different cultural operation to compute P inputs and outputs.

- (a) P/N (nitrogen) ratios for cattle excreta inputs on grazed grassland (Vertès, unpublished data) and spread manures (Bodet et al., 2001)
- (b) amounts of spread mineral P at sowing by crops, depending on the associated N fertilization (Payraudeau et al., 2006)
- (c) P/N ratios for mature crops used to compute harvested P (Schvartz et al., 2005). 0\* signifies that the crop is not harvested but destroyed by plowing (grasslands and catch crops)
- (d) P/DM (dry matter) ratios used to compute grazed and mowed P (Schvartz et al., 2005)

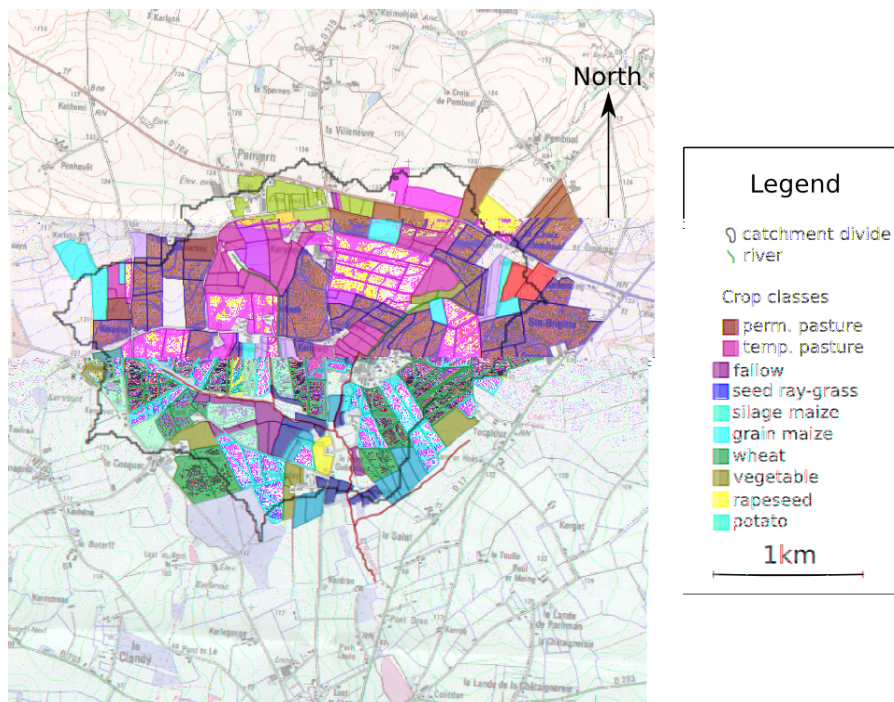


Fig. IV.26: Agricultural landcover of the Kervidy catchment in 2000, for the fields associated to a known farmstead (adapted from Bordenave et al., 2005 and Payraudeau et al., 2006).

## **Factorial simulation design**

A set of virtual landscapes differing by soils, climate, and cropping systems were derived from characteristics and structures observed on a real farming catchment, Kervidy, located in Brittany (Western France). Since the virtual landscapes were not submitted to the same cropping systems and climate constraints, the simulations were designed to last 40 years to allow the systems to stabilize and facilitate the analysis of the results. However, such long records for daily climate series, landcover dynamics and agricultural practices do not exist and therefore were simulated.

### **The Kervidy catchment (Brittany, western France)**

The Kervidy catchment was the case-study site used in the first application of TNT2 (Beaujouan et al., 2002) and has been studied extensively since 1971 (Cheverry, 1998). It covers an area of 5 km<sup>2</sup>, of which 80% is dedicated to intensive farming and breeding (Fig. IV.26). The main farm-types of Kervidy are: pig (216 ha), dairy (86 ha), dairy-pig (72 ha), and other (14 ha, this class groups farmsteads of unknown activity or with marginal farm activities) (Payraudeau et al., 2006). It has a smooth relief and its silty loam soils recover a shallow and impermeable bedrock of Brioverian schist. Between the upper soil and the bedrock, there is a weathered Brioverian schist layer (or regolith) that is several meters thick (from 1 to 30 m).

Several datasets were combined to provide the materials used in the construction of the virtual landscapes. A V-notch weir and a weather station were installed in 1993 which provided a 13-year long series of climate and discharge data on a daily basis. Topographic characteristics were derived from a 20-m digital elevation model (DEM) produced by ISTAR. A detailed soil map at 1/10000 scale was established by Walter and Curmi (1998). A systematic inventory of summer crops was done on-field between 1993 and 2002 by Bordenave et al. (2005) and a life-cycle analysis of farmstead practices was undertaken in 2003 by Payraudeau et al. (2006). A complementary dataset, concerning a neighboring catchment characterized by similar farming activities (the Frémur catchment in Brittany, France), was used to determine winter landcover allocation rules and fertilization practices (Durand et al., 2006).

### **Soil pattern factor levels**

Three soil patterns were constructed by combining distributed maps of soil thickness and waterlogging soil classes. The soil map drawn by Walter and Curmi (1998) involved 487 auger drills scattered over the catchment. The original thematic map of the soil waterlogging classes was derived from the soil map by thresholding the original waterlogging indicator. The soil waterlogging classes are:

- well-drained: no soil redoximorphic feature within the first 80 cm. This class represents 43% of the total area;
- intermediate: weak soil redoximorphic features at less than 80 cm, 44%;
- waterlogged: intense redoximorphic features from the surface, 13%.

The estimated thickness of the weathered schist layer for the well-drained and intermediate soil class was approximated to around 25 m, and 12 m for the waterlogged class.



Soil thickness was measured at the sample sites and was spatialized by conditionally simulating soil thickness with the Gstat package for R (Pebesma, 2004). In the sequential simulation algorithm used by Gstat, the prediction locations are visited in some random sequence. When a certain location is visited, the conditional distribution (conditional to data and previously simulated values) at that location is obtained from the variogram and neighboring samples and a value is drawn randomly from this distribution and added to the conditioning set (Pebesma and Wesseling, 1998). To avoid negative values of simulated thickness, we used a normal score transform (NST) to normalize soil thickness data (Goovaerts, 1997), to fit a spherical variogram using the least square approximation, and finally to revert the simulated values into soil thickness with the inverse NST table. The combination of the original soil waterlogging class map and the simulated thickness map will be referred to as the “observed” soil pattern.

The “thin” soil pattern combined the observed waterlogging class map with soil thickness estimates using the same conditional approach, but by preliminary halving the sampled thickness values before the NST. The resulting map was linearly rescaled to avoid computational issues with TNT2, so that the minimum thickness was 0.3 m.

The “extended waterlogged” soil pattern combined the observed thickness map with the waterlogging soil class map for which the waterlogged class was artificially extended. The downstream topographic index (DTI) provides an efficient way to predict the spatial distribution of soil waterlogging (Chaplot et al., 2000) and is defined by equation 1:

$$DTI = \ln \left( \frac{Area_{upslope}}{DG} \right) \quad (1)$$

where  $Area_{upslope}$  is the upslope contributing area of the location for which DTI is computed, and  $DG$  is the downslope gradient which considers the shortest flow path to the stream. DTI values increase for pixels with larger upslope contributive areas and smaller downslope gradient, which characterize valley bottom wetlands. We managed to double the area occupied by the waterlogged soil class by aggregating pixels with DTI values above 4.

### **Climate factor levels**

Thirteen hydrological years were recorded by the weather station. The “observed” climate level was constructed by artificially looping the 13-year climate series three times. The series of the “dry” climate level was constructed by ordering the hydrological years according to their annual cumulated rainfall and by randomly sampling hydrological years among the six years characterized by the lowest cumulated rainfalls. The series of the “wet” climate level was constructed similarly by sampling among the six years characterized by the highest cumulated rainfall. In absence of measured data, atmospheric N deposition was estimated to 5 kg N.ha<sup>-1</sup>, which is in the range of values measured in other neighboring catchments (Durand et al., 2006). The daily N deposition amount was proportional to the daily rainfall, thus assuming a constant N concentration in rainwaters over a hydrological year.

### **Cropping system factor levels**

Two levels of cropping systems were compared: “intensive” and “moderate”. They differ mainly by winter landcover and nutrient management. The “intensive” and “moderate” levels correspond respectively to the practices observed during the 1995-1997 and 2001-2003 periods on the Frémur catchment. In the time span between these two periods, significant efforts were done by farmers to reduce soil N inputs. For each period, farm-type, and landcover class, a typology of CMP was established based on the origin of the fertilizers involved (mineral, organic) and the spread amounts (Durand et al., 2006). Each landcover class (for a given period and farm-type) was usually characterized by three CMP modalities (up to eight for maize) and their relative proportions were used as probabilities to combine landcover and CMPs in order to produce the data entries for TNT2. Considering fertilization, the “moderate” level differs mainly from the “intensive” level by a reduction of spread mineral and organic N along with a diminution of the grazed surface and excreta inputs to grazed grasslands. Organic P inputs were estimated by the soil P model presented previously. Mineral P inputs for the intensive cropping system were estimated a survey done in 2003 (Payraudeau et al., 2006) from mineral P loads by farm-type and landcover class. For the moderate level, mineral P inputs were systematically omitted.

The agricultural landcover dynamics were simulated at the field scale in two steps:

- Stochastree (Sorel et al., submitted; see also chapter III of this thesis), a model based on stochastic decision trees was used to simulate summer crop transitions by considering temporal (crop rotation) and spatial driving factors (farm-type crop area objectives, spatial distribution of crops around the farmsteads, and sensitivity of crops to soil waterlogging). Stochastree was trained on the 1993-1998 period and the simulation started on year 2000, which was the most exhaustively surveyed year;
- winter landcover was simulated with a set of deterministic and stochastic expert rules, which involved previous and subsequent summer landcover, derived from the Frémur catchment survey (Salmon-Monviola et al., submitted). Allocating winter landcover mainly consisted in managing the harvested maize fields. In the “intensive” level, the fields were left as bare soils generally, whereas in the “moderate” level, they were used to grow catch crops like Italian ray-grass or mustard, which traps N and are then buried by a plowing before the next summer crop sowing (Baggs et al., 2000).

The observed and simulated areas of each crop were compared by a Wilcoxon test ( $p < 0.05$ ) to assess the ability of the crop transition model to maintain the observed stationary proportions. The simulated areas of winter and summer crops were also compared between the “observed” and the “extended waterlogging” soil patterns to ensure that these soil patterns do not differ also in terms of crop areas.

Since the summer crop transition model accounts for crop sensitivity to soil waterlogging, it was necessary to simulate another landcover dataset for the “extended waterlogging” soil component.

<i>variable</i>	<i>unit</i>	<i>description</i>
<b>Model inputs (illustrative in PCA)</b>		
rainfall	mm	cumulated rainfall
NfertiMin	kg N.cultivated ha <sup>-1</sup>	amount of N spread as synthesis fertilizer per cultivated hectare
NfertiOrg	kg N.cultivated ha <sup>-1</sup>	amount of spread N manure and excreta inputs to grazed grassland per cultivated hectare
NOrg/Tot	%	total organic N (NfertiOrg) fraction of total N inputs (NfertiMin plus NfertiOrg)
Pinputs	kg P.cultivated ha <sup>-1</sup>	amount of P inputs (spread mineral and organic fertilizers plus excreta inputs to grazed grassland) per cultivated hectare
<b>Agricultural indicators</b>		
Nharvest	kg N.cultivated ha <sup>-1</sup>	amount of N harvested with crops
Ngraze	kg N.cultivated ha <sup>-1</sup>	amount of N grazed by cattle
Nmow	kg N.cultivated ha <sup>-1</sup>	amount of N mowed from grasslands
Nefficiency	%	agricultural efficiency ratio of N inputs (grazed, harvested, mowed) over inputs(fertilization, excreta inputs to grazed grassland)
Poutputs	kg P.cultivated ha <sup>-1</sup>	total amounts of P outputs (harvested, grazed, mowed)
Pefficiency	%	agricultural efficiency ratio of P inputs over outputs
<b>Soil indicators</b>		
Nmine	kg N.ha <sup>-1</sup>	amount of mineralized N
Ndenit	kg N.ha <sup>-1</sup>	amount of denitrified N
Nawc	kg N.ha <sup>-1</sup>	mean N amount dissolved in the available water content
NsoilBudget	kg N.cultivated ha <sup>-1</sup>	leachable N pressure: N budget between ammoniacal inputs (fertilization and grazing restitution, soil N mineralization) and N inputs (harvests, mows, grazing)
PsoilBudget	kg P.cultivated ha <sup>-1</sup>	soil surface budget between P inputs and outputs
<b>Hydrological indicators</b>		
discharge	mm.day <sup>-1</sup>	catchment specific discharge (water volume divided by the catchment area)
saturatedArea	%	mean fraction of the catchment area for which the groundwater reached the soil surface (soil saturation)
penetratedArea	%	mean fraction of the catchment area for which the groundwater table reached the soil layer (partial soil saturation)
[N]groundwater	mg N.l <sup>-1</sup>	mean groundwater N concentration
[N]stream	mg N.l <sup>-1</sup>	mean stream N concentration at the outlet
[N]streamExcessDuration	day	total duration during which the daily stream N concentration was above the 11.3 mgN.l <sup>-1</sup> EU regulation threshold
[N]streamExcess	mg N.l <sup>-1</sup>	mean stream N concentration in situation of excess
Nstream	kg N	load of N discharged at the outlet
Nstream/fertiTot	%	N transmission ratio between the N load over the amount of N spread as synthesis and organic fertilizers

Table IV.6: Description of active and illustrative variables used in the statistical analyses.

## **TNT2 simulations**

### **TNT2 calibration with a reference scenario**

The so-called “reference virtual landscape”, which combines the “intensive” cropping system, “observed” soils, and “observed” climate levels, was used to calibrate the TNT2 model over the 1994-2005 decade for which we had measured data. The calibration was performed by a trial and error method considering both graphical and mathematical criteria. Nitrogen fate processes of TNT2 rely on water fluxes (infiltration and exfiltration, vertical leaching, surface and sub-surface horizontal fluxes), hydrological parameters related to soil lateral transmissivity and vertical conductivity were adjusted to allow a good reproduction of the stream specific daily discharge, which was assessed by the Nash-Sutcliffe coefficient (Nash and Sutcliffe, 1970) defined by equation 2:

$$E = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - \mu_o)^2} \quad (2)$$

where  $O_i$  and  $P_i$  correspond respectively to the observations and the predictions and  $\mu_o$  is the average of the observations.

Then, nitrogen transformation parameters related to mineralization rate were adjusted to reproduce both fluctuations of measured N concentration and the global mean value. However, the N stream concentration sampling was sparse on occasions and not regular in time. Therefore instead of computing biased Nash-Sutcliffe coefficients, a graphical interpretation was preferred.

### **Initialization of all the virtual landscapes**

Eighteen different virtual landscapes were created by combining each level of the different soil, climate and cropping system factors. Each one served as an input dataset for TNT2 and the soil P model with a common initial state, which was created by a six-year run of TNT2 with the reference scenario to produce distributed maps of soil nutrient contents, and groundwater and soil available water content (AWC) reservoirs and N concentration.

### **Statistical analysis**

Since all landscapes were similarly initialized but then submitted to different climate series and cropping systems, the period of time necessary for the systems to stabilize was ignored when performing the following analyses, which do not expect any temporal trend in the data. Simulation outputs were aggregated annually by either summing fluxes or averaging concentrations and amounts over the whole catchment (Table IV.6) in order to remove seasonal variability. Assuming that landscapes reached stationary state, annual values were considered as independent repetitions and their variance thus reflects interannual variability. Statistical analyses were conducted using the R software (R Development Core Team, 2007).

## Principal component analysis of simulation results

We focused our analysis on three kinds of numeric variables (Table IV.6):

- agricultural indicators: nutrient inputs and use efficiency;
- soil indicators: mineralized and denitrified N amounts, N concentration of AWC, soil N and P budgets;
- hydrological indicators: discharge, stream N concentration, groundwater N concentration, soil saturation rate.

A PCA was done on the aforementioned variables as active variables with the *ade4* R package (Dray and Dufour, 2007), which allowed to project the variables on the PCA factorial plans along with illustrative variables (model entries: rainfall, nutrient inputs) as well as the data values with their relative soil, climate, and cropping system level labels. The latter functionality permits to explore how these factors structure the simulation results.

## Analysis of variance

Each simulation output variable was examined with factorial ANOVA procedures in order to study *(i)* the proportion of variance explained by the factors and interactions *(ii)* the individual effect of the factor and interaction levels. The ANOVA assumptions of data normality and homoscedasticity across factor levels were checked respectively with Shapiro-Wilks (Shapiro and Wilks, 1965) and Bartlett tests (Snedecor and Cochran, 1989) ( $p < 0.05$ ). Since no relative importance was given *a priori* to any factor in our experiment design, a mixed-effect ANOVA model (*ie* model III) was used to analyze the experiment, which tests factor and interaction effect independently of their order in the ANOVA model (Doncaster and Davey, 2007). For each variable, the initial ANOVA model explored the main effects and the pairwise interactions of soil, climate, and cropping system factors. Then, we removed iteratively the model element characterized by the highest p-value if it is above 0.05, recomputed the ANOVA, and this until all the model components were characterized by p-value lower than 0.05. The F-test table of the ANOVA shows how the total variance of the studied variable  $X$  is partitioned in element variances (an element being a factor or an interaction) and residual variance (the variance not explained by the model). The proportion of variance explained by an element is the ratio of the element variance over the total variance. The T-test table of the ANOVA shows how each element level modifies the value of  $X$ .

## Results

### *Virtual landscape construction*

#### **Soil pattern factor levels**

Maps of soil thickness and waterlogging class are presented in Fig. IV.27 and their characteristics are summarized in Table IV.7. The waterlogged class for the “thin” and “observed” soil patterns appeared only in the valley bottomland, whereas the well-drained class occupies the catchment divides and the South-East part of the catchment, which are characterized by steeper slopes. The flatter northern part of the catchment is occupied by the intermediate soil class. The observed waterlogged soil class was spatially extended by incorporating pixels with a DTI higher than 4. The waterlogged area doubled from 14% to 29% of the catchment area, and the added area came mainly from pixels of the intermediate class located in the northern part of the catchment (Fig. IV.27).

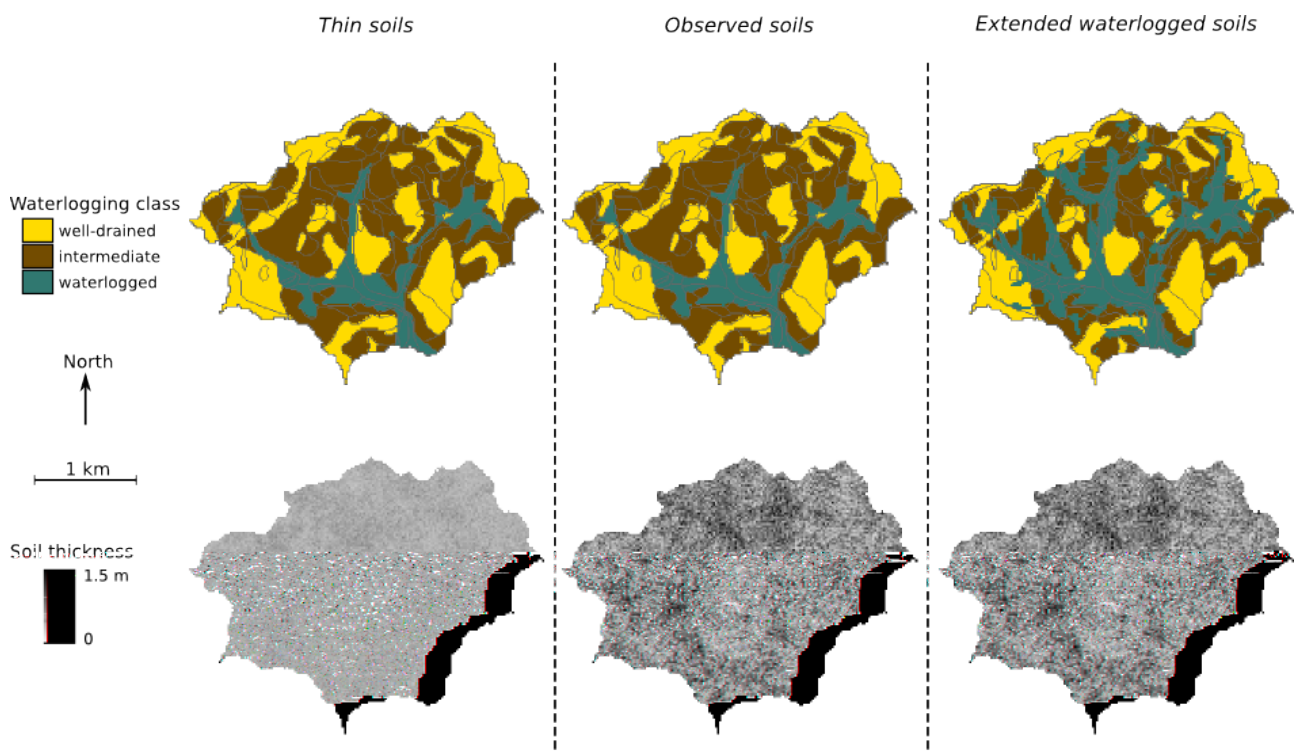


Fig. IV.27: Soil waterlogging class and thickness maps used in the three soil patterns of the virtual landscapes.

	virtual landscape with		
	<i>thin soils</i>	<i>observed soils</i>	<i>extended waterlogged soils</i>
<b><i>waterlogging class proportion</i></b>			
well-drained area	36%	36%	<b>33%</b>
intermediate area	49%	49%	<b>38%</b>
waterlogged area	14%	14%	<b>29%</b>
<b><i>statistics on soil thickness</i></b>			
mean thickness (m)	<b>0.49</b>	0.63	0.63
thickness standard deviation (m)	<b>0.06</b>	0.22	0.22

Table IV.7: Waterlogging class and thickness characteristics of the three soil patterns of the virtual landscapes. Bold values show the specificity of the “thin soils” and “extended waterlogged soils” patterns.

The NST transformations of measured and halved thickness produced the same normalized dataset, therefore the fitted spherical variograms shared the same parameters of equation 3 (fit mean square error = 2.2e-05):

$$\gamma(h) = \begin{cases} 0.71 + 0.25 \left( \frac{3}{2} \left( \frac{h}{424} \right) - \frac{1}{2} \left( \frac{h}{424} \right)^3 \right), & \text{if } 0 \leq h \leq 424 \\ 0.96, & \text{if } h > 424 \end{cases} \quad (3)$$

where  $\gamma(h)$  is the normalized thickness variance (before reverse NST) expected between two points separated by an horizontal distance of  $h$  (in m). The fitted variogram showed a high nugget value (0.71), which represents 74% of the sill variance (0.96), and lead to noisy simulated thickness maps (Fig. IV.27). Although the simulated maps for observed and thin thicknesses differ in range, mean and standard deviation values, the same original spatial structures are still noticeable in both thickness maps: brighter zones, corresponding to thinner soils, can be found at the center of the catchment for instance. The simulated “observed” thickness map (Fig. IV.27) was characterized by a mean thickness slightly lower than the mean sampled thickness (0.63 m instead of 0.69 m, Table IV.7) and the same standard deviation (0.22 m).

### Climate factor levels

Table IV.8 compiles statistics for the generated climate series. The dry climate series was characterized by the lowest mean annual rainfall, and the highest mean potential evapotranspiration (PET) and global radiation; and the opposite for the wet series. The dry climate series is likely to cause frequently crop hydric stresses since mean PET (0.72 m) is higher than the mean annual rainfall (0.66 m). The relatively small differences of mean minimum or maximum temperature among climate series illustrate the mild and temperate climate of Brittany.

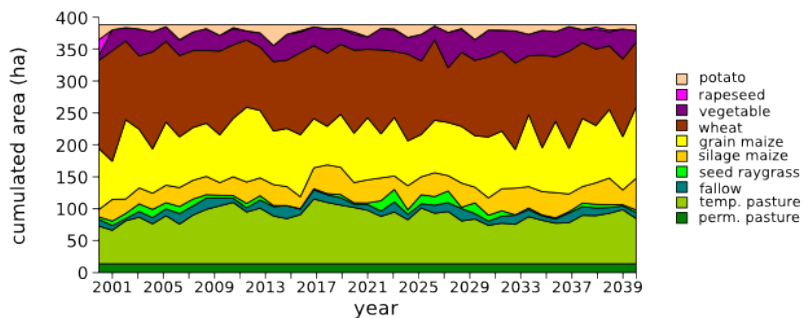
	virtual landscape with		
	<i>dry climate</i>	<i>observed climate</i>	<i>wet climate</i>
<b>characteristics</b>			
annual rainfall (m)	0.66 ± 0.14	0.84 ± 0.25	1.07 ± 0.26
annual PET (m)	0.72 ± 0.11	0.70 ± 0.11	0.69 ± 0.10
minimum temperature (°C)	7.7 ± 0.7	7.8 ± 0.6	8.0 ± 0.5
maximum temperature (°C)	14.9 ± 2.3	15.3 ± 1.5	15.9 ± 0.5
global radiation (MJ.m <sup>2</sup> )	11.9 ± 1.0	11.6 ± 0.9	11.5 ± 0.7

Table IV.8: Climate components of the virtual landscapes (mean value ± standard deviation)



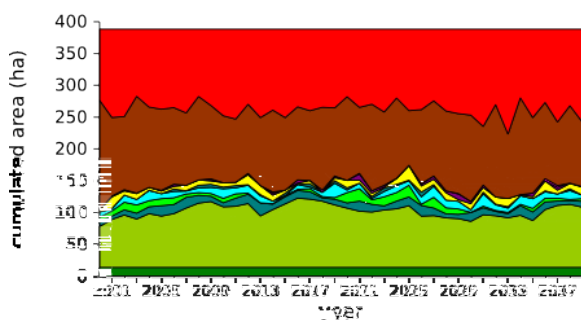
## Crop area evolution

Summer crop dynamics (common to intensive and moderate fertilizations)

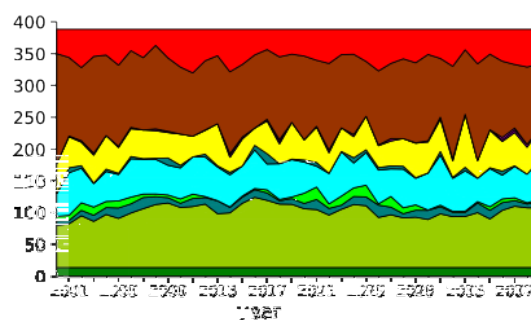


Winter crop dynamics

Intensive practices



Moderate practices



## Nutrient inputs (a: nitrogen, b: phosphorus)

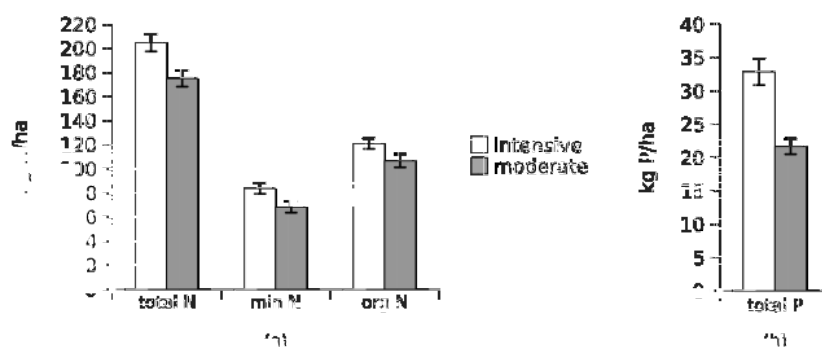


Fig. IV.28: Agricultural components of the virtual landscapes: simulated agricultural landcover dynamics of the Kervidy catchment, annual nutrient input loads by cultivated hectare (means and interannual standard deviations). Nitrogen forms: min N=synthesized mineral N, org N=organic N from manure.

### **Cropping system factor levels**

Fig. IV.28 shows the resulting cultivated area dynamics of summer and winter crops for scenarios with “observed” and “thin” soils, and the mean loads of N and P applied on the crops for each cropping system. Comparison tests with the observed crop areas did not reveal significant differences for marginal crops, permanent pastures, and for wheat. The area dedicated to grain maize was significantly underestimated by 19% whereas silage maize was significantly overestimated by 35%. However, the simulated and observed total maize area were not significantly different. The simulated area dedicated to temporary pastures was significantly overestimated by 26%. Comparison tests between summer and winter crop simulated areas between the “observed” and “extended waterlogging” components revealed no significant differences for each crop.

Winter landcover for the “intensive” cropping system level was characterized by a large area left as 32% of bare soils (32% of the cultivated area), whereas “moderate” level exhibited only 12% of winter bare surfaces. The “intensive” level was characterized by annual inputs of 205 kg N.cultivated ha<sup>-1</sup>.year<sup>-1</sup> and 33 kg P.cultivated ha<sup>-1</sup>.year<sup>-1</sup>, compared to 175 kg N.cultivated ha<sup>-1</sup>.year<sup>-1</sup> and 22 kg P.cultivated ha<sup>-1</sup>.year<sup>-1</sup> for the “moderate” level. The proportion of organic N (spread manures and excreta inputs to grazed grasslands) for both cropping systems was around 60%. The “moderate” level represents a 15% reduction of the mean N input load, and 33% of the mean total P load.

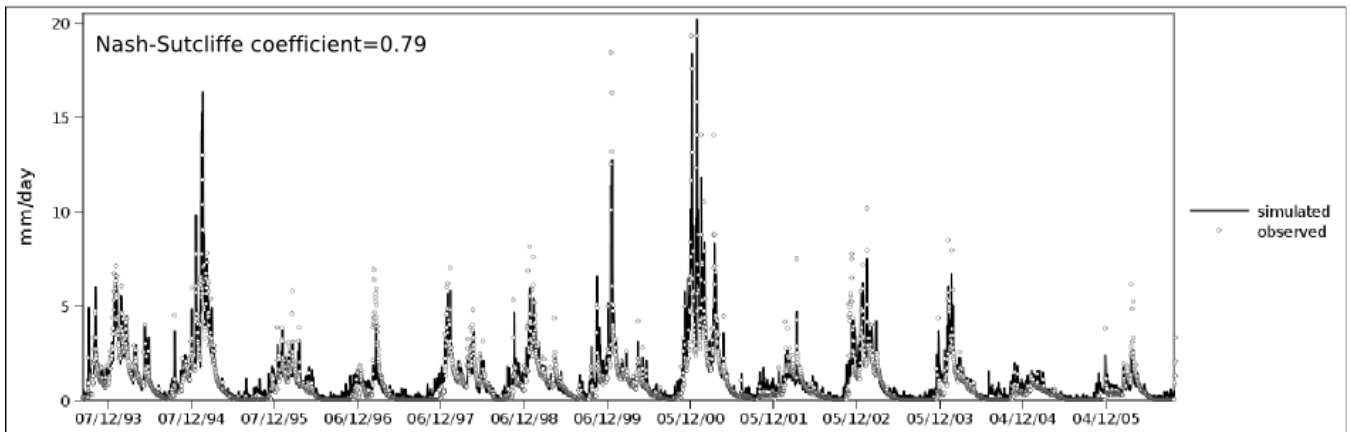
### **Model calibration**

The adjustment of the hydrological parameters allowed a good reproduction of the stream specific daily discharge (Fig. IV.29), with a Nash-Sutcliffe value of 0.79. The global mean value of the simulated stream N concentration differed only by 0.2 mg N.l<sup>-1</sup> with the measured N concentration. Until 2000, N concentration data are sparse and the apparent good reproduction of measurements is therefore hard to validate. During the 2000 civil year, simulated concentrations underestimated the measurements most likely because the “intensive” cropping system also underestimated the corresponding N input load. During the last four years, simulated and measured concentrations matched relatively well, despite the difference between the simulated and observed agricultural landcover.

The mean annual mineralized N amount simulated by TNT2 was 161 kg N.ha<sup>-1</sup>, which is in the higher side of the 90 – 160 kg N.ha<sup>-1</sup> range reported in French soils by Mary and Rémy (1979) and Mary et al. (1996). The mean daily denitrified amount was 29 kg N.ha<sup>-1</sup>.year<sup>-1</sup>.

Concerning P, the mean annual simulated soil P budget equaled 16 kg P.cultivated ha<sup>-1</sup>. Considering that plowing and bioturbation would mix P inputs in an average thickness of 0.4 m and a mean soil bulk density of 1.4, the P budget corresponds approximately to a soil P content increase of 2.85 mg P.soil kg<sup>-1</sup>.year<sup>-1</sup>, which is coherent with the 2.7 mg P.soil kg<sup>-1</sup>.year<sup>-1</sup> rate observed in Brittany between 1980 and 1999 within the French Soil Analysis Database, which gathers more than 8000 soil analyses done in Brittany on agricultural soils (Lemerancier et al., 2006).

### Stream specific discharge



### Nitrogen concentration at the outlet

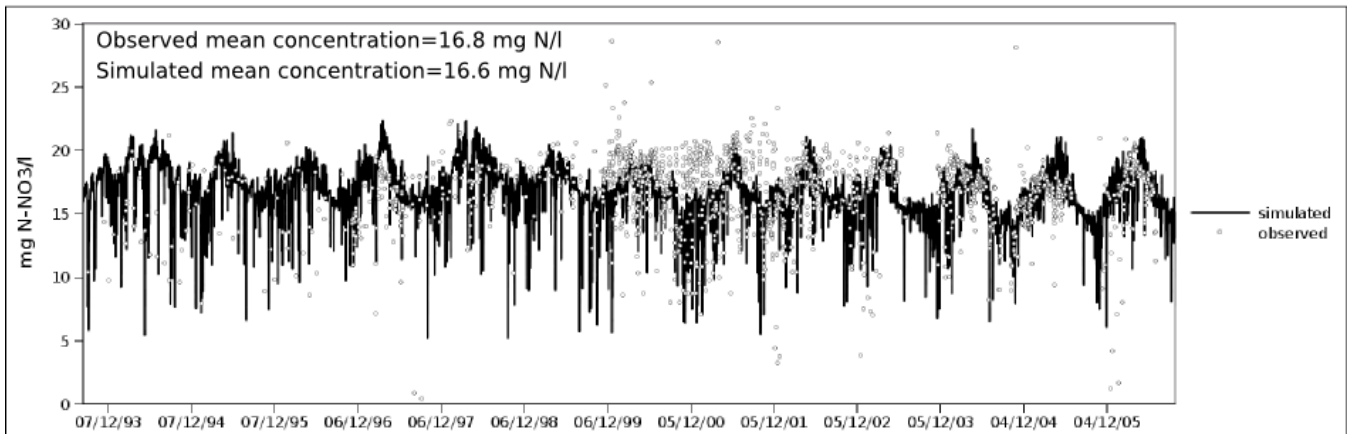


Fig. IV.29: Measured and simulated specific daily discharge and nitrogen concentration at the outlet.

## Scenario comparison and analysis

Statistics of the mean annual N concentrations for the last 17 years of the simulation (the stationary state period) reveal the diversity of responses of each virtual landscape (Fig. IV.30). The simulated outlet N concentration reduced with dryer climate levels, moderate cropping systems, and by switching from the “thin” soil level to the “observed” and then the “extended waterlogged” levels. However, among the four landscapes which have a concentration distribution below the 11.3 mg N.l<sup>-1</sup> regulation threshold only one is characterized by the observed climate level (moderate cropping system, extended waterlogged soils), the three others are characterized by a wet climate level.

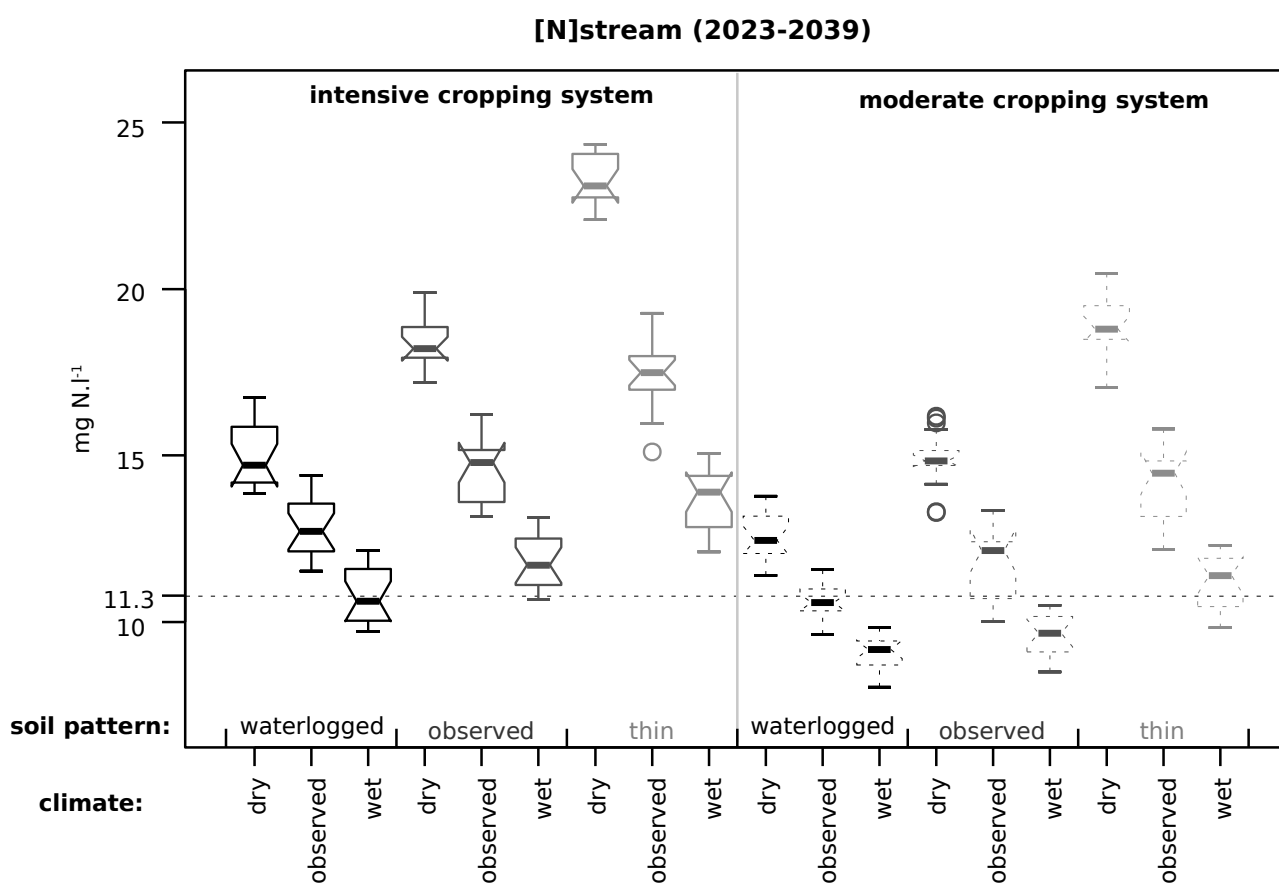
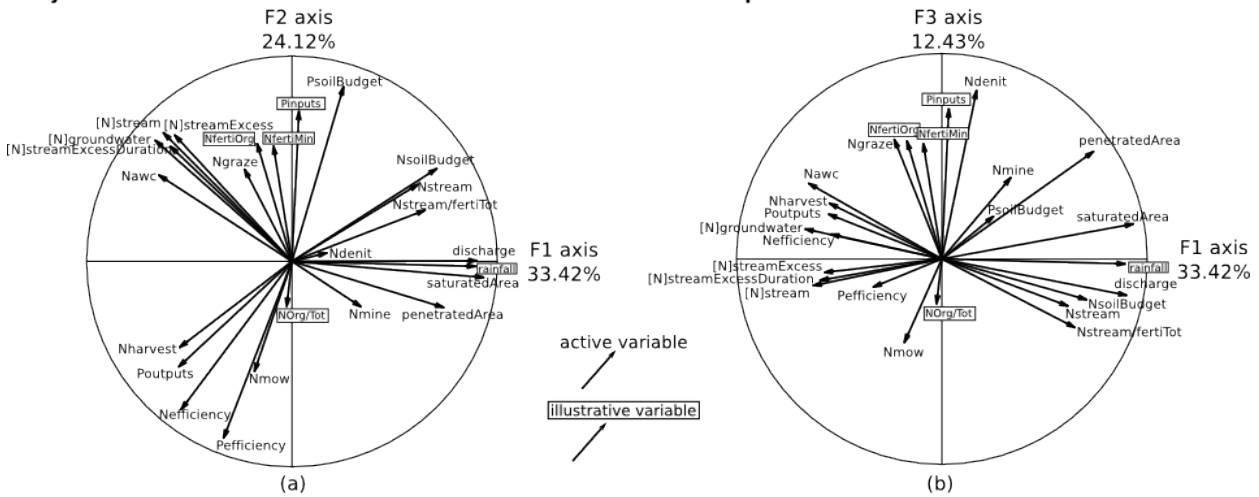


Fig. IV.30: Interannual statistics of stream nitrogen concentration at the outlet for each virtual landscape, over the last 17 years of simulation. The horizontal dotted line represents the European regulation 11.3 mg N.l<sup>-1</sup> threshold (50 mg NO<sub>3</sub>.l<sup>-1</sup>). Black boxplots represent the extended waterlogged soil pattern level, in dark gray the observed level, in light gray the thin level. Plain line boxplots represent the intensive cropping system level, dotted line boxplots represent the moderate level. Climate levels are indicated in the x-axis legend. Meaning of the boxplot shape elements: box basis, middle-stroke and top are 1<sup>st</sup>, 2<sup>nd</sup>, and 3<sup>rd</sup> quartiles; notch=non-parametric 95% confidence intervals for median comparison; lower and higher whiskers=1.5\*interquartile ; circles=outliers.

**Projection of active and illustrative variables on the PCA factorial planes**



**Projection of data, grouped by factor levels, on the PCA planes**

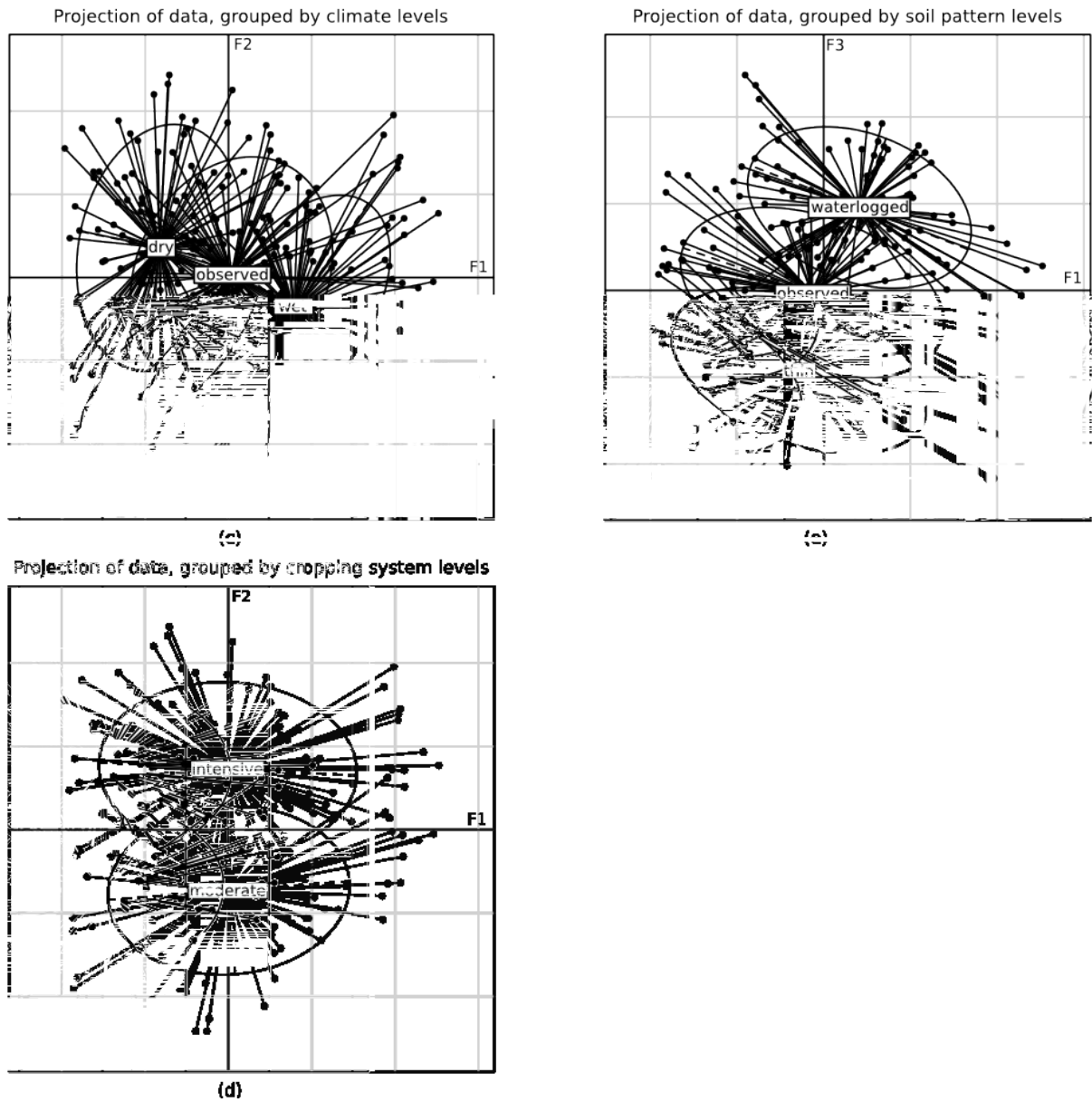


Fig. IV.31: Principal component analysis (PCA) of the results output by TNT2 and the soil phosphorus model. Projection of the variables on the F1-F2 (a) and F1-F3 PCA (b) planes. Projection of the individuals grouped by factor levels: climate levels on F1-F2 (c), cropping system levels on F1-F2 (d), soil pattern levels on F1-F3 (e).

### **Exploratory analysis of the simulation results**

The first three axes of the PCA carried 70% of the variability of the simulation results (Fig. IV.31), 33.4% for axis F1, 24.1% for F2, and 12.4% for F3. The projection of the variables on the F1-F2 factorial plane shows that F1 is positively correlated with hydrological variables such as annual discharge (which is highly correlated with the annual rainfall illustrative variable), the mean annual area proportion of the catchment soil layer reached or saturated by groundwater. F2 is positively correlated with the annual soil P budget (and with illustrative N and P inputs), and negatively with P efficiency ratio. On the first diagonal of the F1-F2 plane, the N soil pressure and the N load at the outlet are negatively correlated with the N efficiency ratio and harvested amounts; on the second diagonal, strong correlations lied between groundwater and stream N concentrations, the total annual duration for which stream N concentration exceeded the regulation 11.3 mg N.l<sup>-1</sup> threshold and the corresponding mean concentration value in excess situation. F3 is highly correlated with denitrified N amounts.

The projection of data values on the factorial planes and their grouping according to soil, climate and cropping system levels show that (i) the centroids of climate groups are aligned with F1 (higher discharges with wetter climates), (ii) cropping system group centroids are aligned with F2 (higher N and P efficiencies with moderate cropping system), (iii) soil pattern group centroids are aligned with F3 (higher denitrified amounts with extended area of waterlogged soils, lower amounts with thin soils).

### **Effects of the cropping system, climate and soil factor levels on model responses**

Normality tests failed for all the studied variables (except mineralized N and the leachable N soil budget) for most of the factor levels (results not shown). Homoscedasticity tests passed more successfully, failing generally for one factor out of three, but failures often occurred for the most explicative factor of the ANOVA model (Table IV.9). ANOVA is known to be more robust for departure from the normality assumption than from homoscedasticity, which modifies the power of the tests (probability to reject the false null hypothesis of ANOVA, that all groups have the same mean value). To normalize the value distribution and homogenize the variance, variable transformation was not considered since it hardly improves the situation for all groups at a time in a multi-factorial analysis. Alternate non-parametric tests equivalent to ANOVA, like Kruskal-Wallis and Friedman tests (Groggel and Skillings, 1986) or rank transformation tests (Conover and Iman, 1981) are respectively inappropriate or little sensitive to factor interaction (Helsel and Hirsch, 2002; Seaman et al., 1994). As a compromise between the ability to study factor interactions and a weakened statistical power, we choose the multi-factorial ANOVA approach and accepted that significant effects may have been rejected.

<i>variables</i>	<i>mean value</i>	<i>percentage of variance</i>				
		<i>cropping system</i>	<i>climate</i>	<i>soil pattern</i>	<i>pairwise interaction</i>	<i>(residual)</i>
<b>Agricultural indicators</b>						
Ngraze	47 kg N.cultivated ha <sup>-1</sup>	43	4	7	1 (cs:s)	45
Nharvest	85 kg N.cultivated ha <sup>-1</sup>		1	9*		90
Nefficiency	73%	21	3	12*		65
Poutputs	16 kg P.cultivated ha <sup>-1</sup>	1	2	7*		91
Pefficiency	62%	67*	1	2		30
<b>Soil indicators</b>						
Ndenit	34 kg N.ha <sup>-1</sup>	2	2	79*	0.3 (s:c), 0.2 (cs:s)	16
Nmine	153 kg N.ha <sup>-1</sup>		5*	18		77
NsoilBudget	87 kg N.cultivated ha <sup>-1</sup>	8	20	3		70
PsoilBudget	11 kg P.cultivated ha <sup>-1</sup>	80*		1		19
<b>Hydrological indicators</b>						
saturatedArea	8%		34*	24		42
penetratedArea	23%		17*	65*	0.8 (s:c)	18
[N]groundwater	16 mg N.l <sup>-1</sup>	12*	76*	1		10
[N]stream	15 mg N.l <sup>-1</sup>	13*	47*	23*	4.3 (s:c), 0.4 (cs:c), 0.3 (cs:s)	12
Nstream	55 kg N.ha <sup>-1</sup>	5*	9*	6		80
Nstream/fertiTot	30%		10*	6		83

Table IV.9: Variance of selected simulation results explained by cropping system, climate, and soil pattern factors (analysis of variance). Outputs are described in Table IV.6. Although significant, null round-off values for variance percentages were removed for legibility reasons (except for pairwise interactions). Asterisks indicate factors for which the hypothesis of homoscedasticity was rejected (Bartlett test,  $p < 0.05$ ). Letters between parentheses identify the pairwise interaction: cs:s=cropping system:soil, s:c=soil:climate, cs:c=cropping system:climate.

The variance of all simulation results could be explained significantly by at least two factors, involving soil and climate in most of the cases (Table IV.9). However the residual variance fraction ranged between around 10% for hydrological indicators like N concentrations, to 90% for indicators related to nutrient outputs. Concerning agricultural indicators, total P outputs, as well as N exported by crops (which represented more than 60% of the total N outputs) have a high residual variance of 90%. The residual variance drops to 45% for grazed N because grazing activities were reduced in the moderate cropping system level. The factor which explains most of the variance for the N and P efficiency ratios is the cropping system (21% and 67% respectively), but the N efficiency ratio was mainly characterized by a high residual variance (65%). The soil indicators were all characterized by a high residual variance (around 75%) except the denitrified N amount and the soil P budget whose variances were explained at 80% by the soil factor (waterlogged soil class is characterized by higher denitrification rates) and the cropping system respectively (the P input load was reduced by one third for the moderate cropping system level). Other hydrological indicators like N load at the outlet and the transmission rate of fertilizers to the outlet had a high residual variance (80%) and half of the remaining variance was explained by the climate factor (10%). The saturation and penetration rates of the water table in the soils are similar indicators, but their variance was controlled by different factors: 34% of the saturation rate variance was explained by climate, whereas 65% of the penetration rate variance was explained by soils. The stream N concentration was the only variable exhibiting an interaction which explained more than 4% of its variance. The N transmission rate between stream N load over total N inputs was not significantly influenced by the cropping system modalities; in both cases, 30% of the N amount applied on fields reached the stream outlet. The N remediation efficiency ratio between the cropping system levels, which is the ratio between the difference of N concentration over the difference of N inputs, equals 0.08. This means that the reduction of the N inputs by  $1 \text{ kg}\cdot\text{ha}^{-1}$  yields a decrease of the N concentration of  $0.08 \text{ mg N}\cdot\text{l}^{-1}$ .



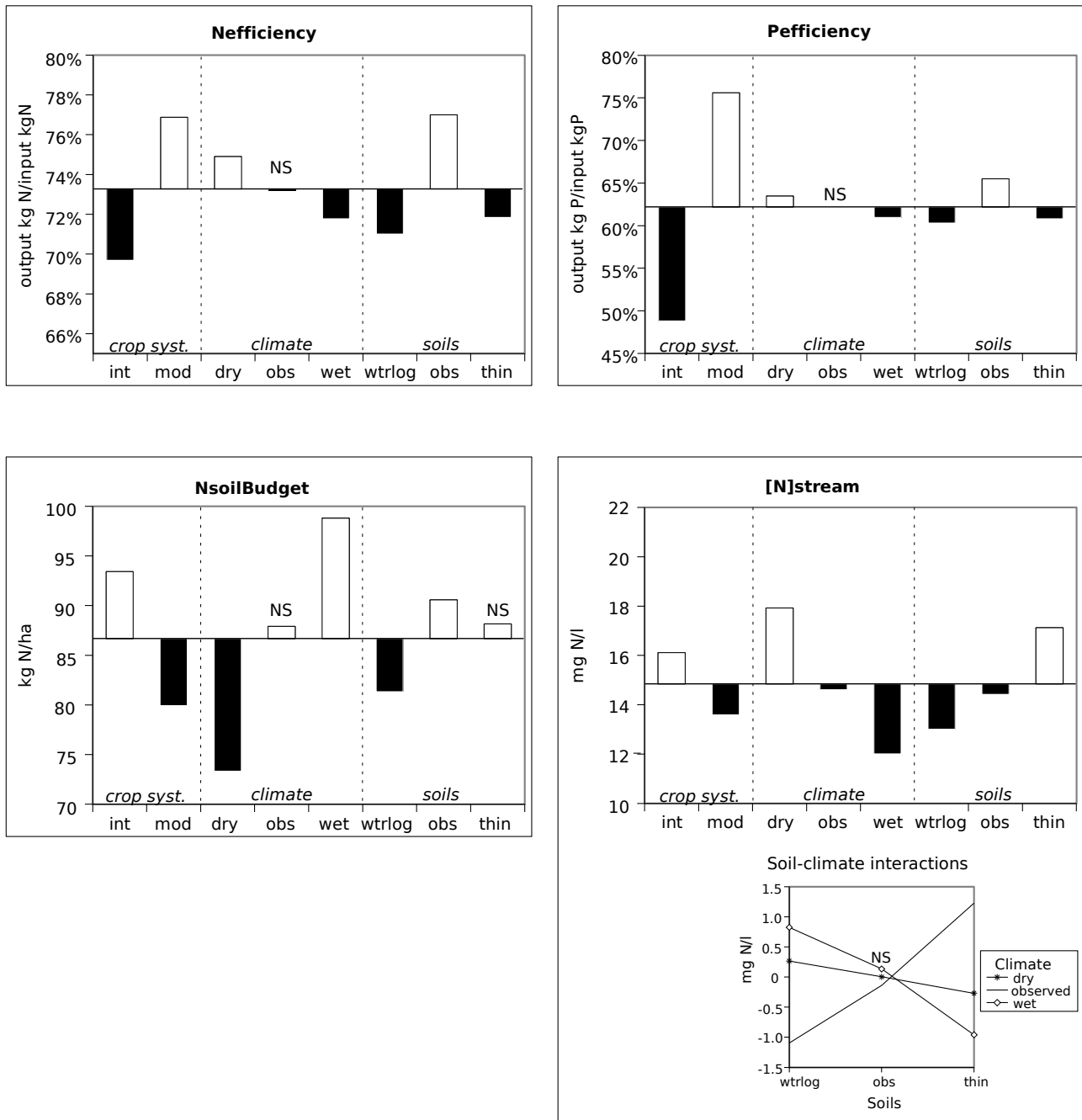


Fig. IV.32: Effects of cropping system, climate and soil pattern levels on selected simulation results computed with an analysis of variance (ANOVA). For a given variable, the horizontal line shows the intercept of the ANOVA model, which is its global mean value. Estimates of the soil-climate interaction are shown for the simulated stream nitrogen concentration. Factor levels are: int=intensive, mod=moderate, obs=observed, and wtrlog=extended waterlogged soil. Non-significant effects are signaled with the tag 'NS'.

The effects of the factor levels varied among the studied variables (Fig. IV.32). Nitrogen and P efficiency ratios were influenced similarly by the factor levels, although P efficiency was mainly sensitive to the cropping system. The efficiency ratios both improved with the moderate cropping system; they decreased with the thin and extended waterlogged levels; and also decreased with wetter climates. The effect of the cropping system and climate factor levels on the N soil budget were opposite than for the N efficiency ratio, but similar for soil pattern levels. The P soil budget presented responses similar to the N soil budget that were not represented on Fig. IV.32. The P soil budget varied mainly with the cropping systems and equaled 17 and 5 kg P.cultivated ha<sup>-1</sup> for the intensive and moderate cropping system respectively. Finally, stream N concentration decreased with a moderate cropping system, wetter climates, and thicker or extended waterlogged soils. However, with the observed climate and soil levels, neither the intensive nor the moderate cropping systems allowed the mean stream N concentration to be below the 11.3 mg N.l<sup>-1</sup> regulation threshold. The magnitude of the significant soil-climate interactions was around 1 mg N.l<sup>-1</sup>, which makes them as important as the magnitude of the cropping system level effects. For the observed climate level (the intermediate climate level), switching from the extended waterlogged soil level to the thin level increased the stream N concentration, which is the opposite behavior than for the dry and wet climate levels. None of the soil-climate interactions involving the observed soil pattern were significant.

## **Discussion**

The use of virtual landscapes allowed the design of a full factorial experiment to compare the effects and the interactions between agricultural, climate, and soil factors on a wide range of multi-resource indicators. Exploratory and statistical analyses permitted to identify the overall controlling factors and to quantify their relative and absolute influences.

### ***Construction of the virtual landscapes***

The creation of the virtual landscapes relied on the combination of initial states (initial landcover, soil thickness and waterlogging classes) with evolution processes (summer crop transition, crop growth and field management), all derived from survey data. While the reference landscape used for calibration was constructed with observed structures, the alternate landscapes were artificial combinations of soil and cropping system components submitted to different climate series. These combinations are obviously in contradiction with the commonly admitted concept in geomorphology that soils are at the interface of the atmosphere, the lithosphere and hydrosphere and serve as a support for human activities, thus making soil morphology a very complex result of interrelated and dynamic processes. However, this approach served our analytical purposes which assumed no hierarchy in the importance of the factor effects. Incorporating correlations between the landscape components would have possibly biased the statistical analyses or complicated the interpretation.

Moreover, in order to ensure the most reliable response of the diffuse nutrient transfer model for each virtual landscape, (i) the model was calibrated on a reference scenario which was composed of the observed

landscape components, (ii) the alternative cropping system, climate and soil components were derived from the observed ones by either sampling specific values (dry and wet climate levels), reducing their range of variation (thin soil level, moderate cropping system level), or modifying the proportion between existing classes (extended waterlogged soil level). Therefore, the alternate virtual landscapes evaluated by TNT2 were comprised in the range of situations encountered during calibration. However, this methodology provides only a limited diversity of new landscapes since they can be considered as “particular cases” drawn from the reference landscape; and as such, they most likely induced the observed heteroscedasticity in simulation outputs.

## ***Multi-resource and multi-criteria analyses***

### **Agricultural efficiency ratios**

The PCA efficiently scattered the simulation results inertia along the first three factorial axes (Fig. IV.31) and provided complementary information which facilitates the understanding of the ANOVA results. On the PCA F1-F2 plan, P and N efficiency ratios appeared strongly negatively correlated with P and N soil budgets respectively. Since (i) higher efficiency ratios were correlated with moderate cropping systems and (ii) N exported by crops variance was not explained by the cropping system factor (Table IV.9), it means that the intensive cropping system level corresponded to a context of over-fertilization, which was not affected by the moderate cropping system level. Reducing the nutrient loads reduced the nutrient budgets and increased the nutrient efficiency ratios (Fig. IV.32), thus the opposite effects of cropping system levels on nutrient budgets and efficiency ratios. A similar phenomenon explains the opposite effects of climate levels on N budget and efficiency ratio: wetter climate levels implied lower N outputs due to lower global yields, thus increasing the N soil budget and decreasing the N efficiency ratio. However the opposite behavior of N budget and efficiency ratio was not observed for the soil pattern effects. The descending order of the soil pattern effects (observed, thin, extended waterlogged) characterizing the N efficiency ratio is due to the same decreasing effects of soil levels on N outputs. For the soil N budget, the same descending order is due to a high mineralization rate in observed soils and a high denitrification rate for extended waterlogged soils.

### **Soil nutrient budgets**

Concerning the N and P soil budgets, the N and P outputs did not decrease with the reduction of the input loads between the intensive and the moderate cropping system levels, meaning that the intensive level implied N and P over-fertilizations. The soil N budget used in this study represents the leachable N pressure remaining after the crop and grassland N outputs, which equaled 93 and 80 kgN.cultivated ha<sup>-1</sup> for the intensive and moderate cropping systems respectively. The fate of this leachable N pressure is either to be denitrified in bottom valley wetlands or discharged at the outlet. Soil characteristics and climate are obvious control factors for denitrification: wetter climates increase the saturated soil rate (10% of the catchment area for the wet climate level, 6% for the dry climate level) and route leached N to saturated soils where denitrification occurs. The waterlogged soil class was characterized by lower water conductivity rates and

higher denitrification rates, thus the mean annual denitrified amount was higher with the extended waterlogged soil level (50 kg N.ha<sup>-1</sup>) than with the observed soil level (26 kg N.ha<sup>-1</sup>).

### **Hydrological indicators**

Dissolved N leaving the catchment through the outlet can be studied either by estimating the load (discharged amount per unit of time) or measuring the concentration. Concentration is often preferred by hydrologists over N load since concentration is easier to measure and less correlated with rainfall or discharge, which was also true in our experiment (discharge and rainfall had the same correlation coefficients of -0.5 with concentration and 0.8 with load). However, Table IV.9 shows that half of N concentration variance is explained by climate (47%), but only 9% for stream N load.

Stream N concentration showed significant soil-climate interactions (except with the observed soil level) with magnitudes similar to the cropping system effects (1 mg N.l<sup>-1</sup>, Fig. IV.32). Denitrification is most likely the main cause of these interactions since it occurs in bottom valley wetlands, just before groundwater reaches the stream and can remove a significant amount of dissolved nitrate (Beaujouan et al., 2002): given the normal climate, denitrification was higher for the extended waterlogged soil level (thus decreasing the stream N concentration) than on the thin soil level because of the higher denitrification rate of the waterlogged soil class. For the thin soil level, the soil volume where denitrification can occur is reduced, and so is the amount of denitrified N (thus increasing the stream N concentration). Given the extended waterlogged soil level, a wetter climate would increase the turnover of infiltrated water whose dissolved oxygen inhibits denitrification. With a dryer climate, the saturation rate of the catchment area decreases from 10% (observed climate) to 12%, thus reducing the soil extent where denitrification can occur. This would explain why both dry and wet climate levels induced lower denitrification rates with the extended waterlogged soil pattern than the normal climate level. Given the thin soil pattern, the wet climate increased the saturation rate of the catchment area from 6% (observed climate) to 9%, thus increasing the denitrification soil extent. But for a dry climate, the saturation rate of the catchment reduces from 6% (observed climate) to 4%, which should imply a reduction of denitrification and a positive effect on stream N concentration instead of the observed negative effect.

### **Remediation practice assessment**

The study involved two nutrients: phosphorus, and nitrogen. The soil P budget was significantly improved from 17 to 5 kg P.cultivated ha<sup>-1</sup> mainly by omitting every synthetic P fertilizer and relying on organic P inputs. However, such a remediation is only possible for farmsteads that produce their own manure or have an easy access to an external source. Similarly, the leachable N pressure applied on soils could be significantly reduced by adopting the moderate cropping system level. However, although both P and N efficiency ratios improved, one third of the N inputs were still routed to the outlet and the impact of N input reduction was really low on the stream N concentration. This highlights the complexity of N fate, both in terms of transformation and fluxes. Moreover, a reduction of the mean soil thickness by 14 cm (22%)

increased the mean stream N concentration by 2 mg N.l<sup>-1</sup>. which add evidence to the importance of soil thickness in N retention (Ruiz et al., 2002; Oehler et al., 2005).

### ***Methodology improvements and perspectives***

For comparison purposes, we aggregated all the information at the landscape level but finer aggregation would be of interest in order to compare agricultural indicators between farm-types or landcover class. Nitrogen and P efficiency have improved at the landscape scale by adopting a moderate cropping system, but it may not be the case at the field scale, when manure spreading is not possible due to steep slopes or waterlogged soils.

Our analytical approach constrained the consideration of variables not submitted to temporal trend only, which we achieved by either waiting for the system to stabilize or to study annual budgets instead of stocks. We studied a relatively small catchment but it took more than 20 years to stabilize. Catchments of larger extents or with deeper soils are likely to have longer hydrological response time. Therefore the duration of the simulation needs to be extended, while making sure that no artifact trend appears in climate series, landcover dynamics, or associated agricultural practices. However, Aspinall (2004) and Pocewicz et al. (2008) showed that landcover and farm management drivers change every couple of decades due to socio-economic changes which can occur either locally (reallocation of fields among the farmsteads) or globally (changes in practice regulations or in the market equilibriums). As this suggests that long and constant simulations are not realistic from a socio-economical point of view, it also signifies that only few larger agricultural catchments may reach the equilibrium state corresponding to the agricultural practices occurring in it.

Perspectives can be considered either by extending the method to new resources or by modifying modeling approaches. Concerning soil resources and processes, tillage practices are not accounted for the moment in TNT2, although it has been shown that reduced tillage and winter catch crops improve soil organic carbon sequestration, structural stability, reduce N and P mineralization rates, subsequent runoff and leaching, and production of nitrous oxide (Kingery et al., 1996; West and Post, 2002; Franzluebbers et al., 2007; Liu et al., 2007). Soil carbon content is also a critical indicator of soil biological and structural qualities (Kanchikerimath and Singh, 2001; Le Bissonnais and Arrouays, 1997). However, TNT2 was initially designed as a N fate model at the catchment scale, mainly aiming at simulating soil-crop interactions and water fluxes. Therefore, the current implementation of organic matter transformation processes does not allow the use of TNT2 neither to simulate carbon dynamics nor to assess related impacts based on them. The assessment of greenhouse gases (GHG) emissions from a farming catchment poses the challenge of accounting for the spatial heterogeneity of the denitrification processes (Oehler et al. 2007) and the ammonia volatilization from cattle excreta which are scattered among the grazed grasslands (Hutchings et al., 2007). Virtual landscapes similar to the ones used in this study could be used as spatialized entry dataset for GHG emission and transport models. The TNT2 developer team is currently undertaking new developments in order to monitor nitrogen gas emissions and fluxes.

## Conclusion

The study develops innovative approaches and results for the multi-resource and multi-criteria assessment of environmental impacts of agricultural practices. We presented a methodology to derive agricultural virtual landscapes from observed data and processes by framing the range of variation of the original landscape features. By comparing the intensive and the moderate cropping systems, we showed that the change of practices done by the farmers between the 1995-1997 and 2001-2003 period did improve the soil P budgets and stream N concentration, but not enough to pass below the EU regulation level of 11.3 mg N.l<sup>-1</sup>. However, N concentrations (stream and groundwater) were less sensitive to practice change than climate change (mean annual rainfalls differing by 0.2 m). Similarly, an extension of the waterlogged soil class in bottom wetland valley reduced the N stream concentration more significantly than agricultural practices. Moreover, interaction between soils and climate may reduce the effect of BMP with a wetter climate occurring on waterlogged soils.

The study showed that although agricultural or hydrological indicators may serve their evaluation purposes individually, understanding the fate of diffuse elements like nitrogen requires their combination since transfer and transformation processes are interrelated. Hydrological fluxes of C and P were neglected in this study although there are strong evidences that they also impact water quality. Phosphorus and dissolved organic carbon models are being developed and similar multi-factorial experiments based on virtual landscapes may provide a benchmark for hypothesis test. More generally, multi-factorial virtual landscapes may be used to perform sensitivity analyses for all kinds of spatially distributed models.

## Acknowledgments

The PhD of the first author was funded by the Région Bretagne and INRA. We are grateful to Safya Menasseri, Fabien Ferchaud, Jordy Salmon-Monviola, and Laurence Lebouille (Chambre d'Agriculture du Morbihan) for stimulating discussions and their insights on the agronomic data of Kervidy and Frémur catchments.

## References

- Aarts H.F.M., B. Habekotté, H. van Keulen. 2000. Phosphorus (P) management in the 'De Marke' dairy farming system. *Nutrient Cycling in Agroecosystems*, 56 (3):219-229.
- Aspinall R. 2004. Modelling land use change with generalized linear models – a multi-model analysis of change between 1860 and 2000 in Gallatin Valley, Montana. *Journal of Environmental Management*, 72:91-103.
- Baggs E.M., C.A. Watson, R.M. Rees. 2000. The fate of nitrogen from incorporated cover crop and green manure residues. *Nutrient Cycling in Agroecosystems*, 56 (2):153-163.
- Basset-Mens C., L. Anibar, P. Durand, H.M.G. van der Werf. 2006. Spatialised fate factors for nitrate in catchments: Modelling approach and implication for LCA results. *Science of The Total Environment*, 367 (1):367-382.
- Beaujouan V., P. Durand, L. Ruiz. 2001. Modelling the effect of the spatial distribution of agricultural practices on nitrogen fluxes in rural catchments. *Ecological Modelling*, 137:93-105.
- Beaujouan V., P. Durand, L. Ruiz, P. Arousseau, G. Cotteret. 2002. An hydrological model dedicated to topography-based simulation of nitrogen transfer and transformation: rationale and application to the geomorphology – denitrification relationship. *Hydrological Processes*, 16:493–507.
- Beven K., Kirkby M. 1979. A physically based variable contributing area model of basin hydrology. *Hydrological Sciences Bulletin*, 24 (1):43-69.
- Beven K. 2002. Towards a coherent philosophy for modelling the environment. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 458 (2026):2465-2484.
- Bodet J.M., S. Hacala, C. Aubert, C. Texier. 2001. Fertiliser avec les engrais de ferme. Institut de l'Élevage, ITAVI, ITCF, ITP.
- Bordenave P., F. Oehler, T. Bioteau, N. Turpin, P. Serrand, P. Saint-Cast, E. Le Saos. 2005. Distributed surface-groundwater coupled model applied to long term water management impacts at basin scale. *Ingénieries*, 44:13-36.
- Brazier R.E., A.L. Heathwaite, S. Liu. 2005. Scaling issues relating to phosphorus transfer from land to water in agricultural catchments. *Journal of Hydrology*, 304:330-342.
- Brisson N., Mary B., Ripoche D., Jeuffroy MH., Ruget F., Nicoulaud B., Gate P., Devienne-Barret F., Antonioletti R., Durr C., Richard G., Beaudoin N., Recous S., Tayot X., Plenet D., Cellier P., Machet J-M., Meynard J-M., Delécolle R. 1998. STICS: a generic model for the simulation of crops and their water and nitrogen balances. I. Theory and parameterization applied to wheat and corn. *Agronomie*, 18: 311-346.
- Carpenter S.R., N.F. Caraco, D.L. Correll, R.W. Howarth, A.N. Sharpley, V.H. Smith. 1998. Nonpoint pollution of surface waters with phosphorus and nitrogen. *Ecological Applications*, 8 (3):559-568.
- Chaplot V., C. Walter, P. Curmi. 2000. Improving soil hydromorphy prediction according to DEM resolution and available pedological data. *Geoderma*, 97 (3-4):405-422.
- Cheverry C. (ed.). 1998. *Agriculture intensive et qualité des eaux*. Science Update, Paris. 297 pages.
- Conover W.J. and R.L. Iman. 1981. Rank transformation as a bridge between parametric and nonparametric statistics: *The American Statistician*, 35 (3):124-129.
- Deasy C., A.L. Heathwaite, R.E. Brazier. A field methodology for quantifying phosphorus transfer and delivery to streams in first order agricultural catchments. 2008. *Journal of Hydrology*, 350 (3-4):329-338.
- Doncaster, C.P. and A.J.H. Davey. 2007. *Analysis of variance and covariance. How to choose and construct models for the life sciences*. Cambridge University Press, Cambridge. 302 pages.
- Dray S. and A.B. Dufour. 2007. The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software*, 22(4):1-20.
- Durand P., F. Ferchaud, J. Salmon-Monviola, F. Goetschel, C. Martin. 2006. Étude sur l'évolution des paramètres nitrates dans les eaux brutes des bassins versants Bretagne Eau Pure et des autres bassins versants bretons. Rapport final, volets 3 et 4. INRA, Bretagne Eau Pure.
- European Commission. 1991. Directive of the Council of 12 December 1991 concerning the protection of waters against pollution caused by nitrates from agricultural sources (91/676/EEC). European Commission, Brussels.
- Franzluebbers A.J., H.H. Schomberg, D.M. Endale. 2007. Surface-soil responses to paraplowing of long-term no-tillage cropland in the Southern Piedmont USA. *Soil and Tillage Research*, 96 (1-2):303-315.

- Gabrielle B., B. Mary, R. Roche, P. Smith, G. Gosse. 2002. Simulation of carbon and nitrogen dynamics in arable soils: a comparison of approaches. *European Journal of Agronomy*, 18:107-120.
- Goovaerts P. 1997. *Geostatistics for natural resources evaluation*. Oxford University Press, New York.
- Groggel D.J. and J.H. Skillings. 1986. Distribution-free tests for main effects in multifactor designs. *The American Statistician*, 40:99-102.
- Gustafson E.J. and T.R. Crow. 1998. Simulating spatial and temporal context of forest management using hypothetical landscapes. *Environmental Management*, 22 (5):777-787.
- Hart M.R., B.F. Quin, M.L. Nguyen. 2004. Phosphorus runoff from agricultural land and direct fertilizer effects: a review. *Journal of Environmental Quality*, 33 (6):1954-1972.
- Heathwaite L., A.N. Sharpley, M. Bechmann. 2003. The conceptual basis for a decision support framework to assess the risk of phosphorus loss at the field scale across Europe. *Soil Science and Plant Nutrition*, 166 (4):447-458.
- Helsel D.R. and R.M. Hirsch. 2002. *Statistical methods in water resources*. USGS, Techniques of Water-Resources Investigations Book 4, Chapter A3.
- Hénault C. and J.C. Germon. 2000. NEMIS, a predictive model of denitrification on the field scale. *European Journal of Soil Science*, 51:257-270.
- Hutchings N.J., J.E. Olesen, B.M. Petersen, J. Berntsen. 2007. Modelling spatial heterogeneity in grazed grassland and its effects on nitrogen cycling and greenhouse gas emissions. *Agriculture, Ecosystems & Environment*, 121:153-163.
- Kanchikerimath M. and D. Singh. 2001. Soil organic matter and biological properties after 26 years of maize-wheat-cowpea cropping as affected by manure and fertilization in a Cambisol in semiarid region of India. *Agriculture, Ecosystems & Environment*, 86 (2):155-162.
- Karlen D.L., E.G. Hurley, S.S. Andrews, C.A. Cambardella, D.W. Meek, M.D. Duffy, A.P. Mallarino. 2006. Crop rotation effects on soil quality at three northern corn/soybean belt locations. *Agronomy Journal*, 98:484-495.
- Karpinets T. V., D.J. Greenwood, J. T. Ammons. 2004. Predictive mechanistic model of soil phosphorus dynamics with readily available inputs. *Soil Science Society of America Journal*, 68 (2):644-653.
- Kingery W.L., C.W. Wood, J.C. Williams. 1996. Tillage and amendment effects on soil carbon and nitrogen mineralization and phosphorus release. *Soil and Tillage Research*, 37 (4):239-250.
- Lal R. 2004. Soil carbon sequestration to mitigate climate change. *Geoderma*, 123 (1-2):1-22.
- Le Bissonnais Y. and D. Arrouays. 1997. Aggregate stability and assessment of soil crustability and erodibility: II. Application to humic loamy soils with various organic carbon contents. *European Journal of Soil Science*, 48 (1):39-48.
- Lemercier B., C. Walter, C. Schwartz, N. Saby, D. Arrouays, S. Follain. 2006. Organic carbon and extractable phosphorus contents monitoring in cultivated soils of three French regions. *Étude et Gestion des Sols*, 73 (3):165-179.
- Lewis D.R. and M.B. McGehan. 2002. A review of field scale phosphorus dynamic models. *Biosystems Engineering*, 82 (4):359-380.
- Liu X.J., A.R. Mosier, A.D. Halvorson, C.A. Reule, F.S. Zhang. 2007. Dinitrogen and N<sub>2</sub>O emissions in arable soils: Effect of tillage, N source and soil moisture. *Soil Biology and Biochemistry*, 39 (9):2362-2370.
- Mary B. and J. Rémy. 1979. Essai d'appréciation de la capacité de minéralisation de l'azote des sols de grande culture. I. Signification des cinétiques de minéralisation de la matière organique humifiée. *Annales Agronomiques*, 30:513-527.
- Mary B., S. Recous, S. Darwis, D Robin. 1996. Interactions between decomposition of plant residues and nitrogen cycling in soil. *Plant Soil*, 181-71-82.
- Mignolet C., C. Schott, M. Benoit. 2007. Spatial dynamics of farming practices in the Seine basin: Methods for agronomic approaches on a regional scale. *Science of The Total Environment*, 375 (1-3):13-32.
- Molénat J., C. Gascuel-Oudou. 2002. Modelling flow and nitrate transport in groundwater for the prediction of water travel times and of consequences of land use evolution on water quality. *Hydrological Processes*, 16 (2):479-492.
- Nash, J.E. and J.V. Sutcliffe. 1970. River flow forecasting through conceptual models. *Journal of Hydrology*, 10:282-290.



- Oehler F., P. Bordenave, P. Durand. 2007. Variations of denitrification in a farming catchment area. *Agriculture, Ecosystems & Environment*, 120 (2-4):313-324.
- Pauwels H., P. Lachassagne, P. Bordenave, J.C. Foucher, A. Martelat. 2001. Temporal variability of nitrate concentration in a schist aquifer and transfer to surface waters. *Applied Geochemistry*, 16:583-596.
- Payraudeau S. and H.M.G. van der Werf. 2005. Environmental impact assessment for a farming region: a review of methods. *Agriculture, Ecosystems & Environment*, 107 (1):1-19.
- Payraudeau S., H.M.G. Van Der Werf, F. Vertes. 2006. Evaluation of an operational method for the estimation of emissions of nitrogen compounds for a group of farms. *International Journal of Agricultural Resources, Governance and Ecology*, 5 (2-3):224-246.
- Pebesma, E.J. 2004. Multivariable geostatistics in S: the gstat package. *Computers & Geosciences*, 30 (7):683-691.
- Pebesma E.J. and Wesseling C.G. 1998. Gstat: a program for geostatistical modelling, prediction and simulation. *Computers & Geosciences*, 24 (1):17-31.
- Pocewicz A., M. Nielsen-Pincus, C. Goldberg, M. Johnson, P. Morgan, J. Force, L. Waits, L. Vierling. 2008. Predicting land use change: comparison of models based on landowner surveys and historical land cover trends. *Landscape Ecology*, 23 (2):195-210.
- Préfecture d'Ille et Vilaine. 2005. Arrêté du 23 novembre 2005 relatif au troisième programme d'action à mettre en oeuvre en vue de la protection des eaux contre la pollution par les nitrates d'origine agricole. 58 pages.
- R Development Core Team. 2007. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.r-project.org>.
- Ruiz L., S. Abiven, P. Durand, C. Martin, F. Vertès, V. Beaujouan. 2002. Effect on nitrate concentration in stream water of agricultural practices in small catchments in Brittany: I. Annual nitrogen budgets, *Hydrology and Earth System Sciences*, 6 (3):497-506.
- Ruiz L., P. Arousseau, J. Baudry, V. Beaujouan, P. Cellier, P. Curmi, P. Durand, C. Gascuel-Oudou, P. Leterme, J.L. Peyraud, C. Thenail, C. Walter. 2006. Conception de bassins-versants virtuels : vers un outil pour l'étude de l'influence de l'organisation spatiale de l'activité agricole et du milieu physique sur les flux d'azote dans les bassins-versants. *In Monestiez, P., S. Lardon, B. Seguin. Organisation spatiale des activités agricoles et processus environnementaux. Update Sciences & technologies, INRA, Paris:337-354.*
- Russel E.W. 1973. Soil condition and plant growth. 10<sup>th</sup> ed. Longmans, London.
- Salmon-Monviola J., Durand P., Ferchaud F., Sorel L. 2008. Land cover and cropping system generation: a modelling tool to construct scenarios as input to a distributed nitrate transfer model. Submitted to *Environmental Modelling & Software*.
- Seaman J.W., S.C. Walls, S.E. Wise, R.G Jaeger. 1994. Caveat emptor: rank transform methods and interaction. *Trends in Ecology & Evolution*, 9:261-263.
- Sharpley A.N., B. Foy, P.J.A. Withers. 2000. Practical and innovative measures for the control of agricultural phosphorus losses to water: an overview. *Journal of Environmental Quality*, 29:1-9.
- Shapiro S.S. and Wilks M.B. 1965. An analysis of variance test for normality. *Biometrika*, 52 (3-4):591-611.
- Schvartz C., J.C. Muller, J. Decroux. 2005. Guide de la fertilisation raisonnée. France Agricole (ed.).
- Snedecor G.W. and Cochran W.G. 1989. *Statistical Methods*, Eighth Edition, Iowa State University Press.
- Sonneveld M.P.W., J.M. Schoorl, A. Veldkamp. 2006. Mapping hydrological pathways of phosphorus transfer in apparently homogeneous landscapes using a high-resolution DEM. *Geoderma*, 133 (1-2):32-42.
- Vadas P.A., W.J. Gburek, A.N. Sharpley, P.J.A. Kleinman, P.A. Moore, M.L. Cabrera, R.D. Harmel. 2007. A model for phosphorus transformation and runoff loss for surface-applied manures. *Journal of Environmental Quality*, 36 (2):324-332.
- Walter C. and P. Curmi. 1998. Les sols du bassin versant du Coët-Dan : organisation, variabilité spatiale et cartographie. *In Cheverry C. (ed.). Agriculture intensive et qualité des eaux. Science Update, Paris.*
- Walter C., R. Rossel, A. Viscarra, A.B. McBratney. 2003. Spatio-temporal simulation of the field-scale evolution of organic carbon over the landscape. *Soil Science Society America Journal*, 67:1477-1486.
- West T.O. and W.M. Post. 2002. Soil organic carbon sequestration rates by tillage and crop rotation: a global data analysis. *Soil Science Society of America Journal*, 66 (6):1930-1946.

## **Chapitre V**

### Conclusion générale

*« Science is a process of ruling out ideas that are not true, always leaving some uncertainty about the ideas we think are true. »*

**Eric J. Gustafson**

*« ± Vous avez du mal à supporter les critiques ?  
± Non. Sauf, évidemment, si elles émanent d'un parent, d'un ami, d'une connaissance ou d'un inconnu. »*

**Franklin P. Jones**

# Chapitre V : Conclusion générale

L'évaluation environnementale des agrosystèmes est un enjeu majeur pour le développement de pratiques agricoles durables, la préservation ou la restauration du bon état écologique des ressources naturelles. L'objectif de ce projet était de proposer une méthodologie d'évaluation environnementale qui soit spatialisée et dynamique afin d'intégrer les interactions entre les effets du milieu physique, du climat et des pratiques agricoles. L'approche suivie dans ce projet s'est basée sur une modélisation du paysage agricole, qui a servi de support conceptuel commun aux différentes étapes de ce travail, à savoir :

- l'analyse des successions culturales observées afin d'identifier les déterminants spatiaux et temporels intervenant dans les transitions de cultures ;
- la construction d'un jeu de paysages agricoles virtuels différant quant à leur milieu physique et les forçages climatiques et les pratiques agricoles auxquelles ils sont soumis ;
- la réalisation d'une évaluation environnementale multi-cibles et multi-critères de ces agrosystèmes en permettant le couplage et l'intégration des sorties entre le modèle agro-hydrologique TNT2 et un modèle de bilan de phosphore.

## Synthèse des résultats

### *Méthodologie générale*

L'état des connaissances concernant la modélisation intégrée et fonctionnelle des paysages et des pratiques agricoles a permis de dégager plusieurs points qui ont servi de cadre à ce travail. La revue des méthodes d'évaluation environnementale des agrosystèmes a identifié trois facteurs majeurs pouvant influencer les impacts des agrosystèmes : le climat (à l'origine des flux d'eau), le milieu physique (support des cultures et des flux) et le système de culture (déterminant l'occupation du sol et les apports d'intrants). Les travaux relatifs à l'organisation des exploitations agricoles ont montré une représentation hiérarchisée des composants agricoles du paysage : les parcelles composent les unités de gestion des exploitations agricoles, celles-ci pouvant être regroupées par types d'activité aux objectifs de production et modes de gestion communs. Quant à la description du milieu physique, dont on ne peut connaître exhaustivement les structures internes, nous avons employé des méthodes stochastiques servant habituellement à représenter cette incertitude ou cette variabilité spatiale, pour simuler des variantes virtuelles de paysages réels.

Par contre eudgP iā e

environnementale nous a permis de choisir un jeu d'indicateurs sur les pratiques agricoles ainsi que sur les ressources sol et eau pour notre étude.

La plateforme *Qualscape* a été développée pour coordonner les étapes de ce projet et regrouper les outils construits pour le mener. Les paragraphes suivants présentent une synthèse des résultats majeurs obtenus au cours de ce projet.

### ***Déterminants spatiaux et temporels de l'évolution de l'occupation du sol***

Notre analyse des successions culturales observées sur Naizin a permis de confirmer la forte influence du précédent cultural comme déterminant temporel de la culture à venir, et de mettre en évidence différents niveaux de déterminants spatiaux : l'influence locale de l'hydromorphie de la parcelle, la distance entre la parcelle et son corps d'exploitation, et le type d'activité principal de l'exploitation (lait, porcs, mixte). Nous avons ainsi pu caractériser la contrainte posée par l'hydromorphie sur la conduite des cultures, ainsi que la spécificité de chaque type d'activité en termes d'objectifs de production (exprimés sous forme de proportions de la surface du parcellaire dédiée à chaque culture) et de structures spatiales des cultures (exprimées sous forme de distances moyennes des parcelles au corps d'exploitation).

Nous avons proposé une méthode innovante pour intégrer ces déterminants dans un outil de simulation des transitions de cultures (*Stochastree*), en construisant par apprentissage des arbres de décision à l'aide de l'algorithme C4.5 (Quinlan, 1993) dont nous avons rendu les feuilles stochastiques. Les résultats de simulation de *Stochastree* ont été confrontés à ceux d'un modèle n'implémentant que le déterminant temporel sous la forme de matrices de probabilités de transition de culture (*Rotomatrix*). Les arbres de décision et les matrices ont révélé une structure commune en ceci que le premier déterminant testé dans les arbres de décision fût systématiquement la culture précédente. Les performances similaires des prédictions de *Stochastree* et de *Rotomatrix* réalisées sur un jeu de données d'évaluation suggèrent que certains déterminants sont implicitement intégrés dans les matrices, par exemple dans le fait que les probabilités observées pour qu'une culture sensible à l'hydromorphie succède à une culture résistante sont faibles. Néanmoins, les simulations long terme ont révélé une supériorité significative de *Stochastree* dans le maintien de la distribution spatiale des cultures autour des corps d'exploitation.

### ***Évaluation environnementale des agrosystèmes***

L'évaluation environnementale des 18 paysages virtuels construits a permis de mettre en évidence l'influence du climat, du milieu et des systèmes de culture sur les différents indicateurs utilisés. Nous avons constaté que l'adoption d'un système de culture modéré (réduction intrants, inter-cultures hivernales) réduit significativement les flux excédentaires de phosphore vers les sols et accroît l'efficacité des apports d'azote et de phosphore. Par contre, cet effet s'atténue pour des indicateurs incorporant des processus de transformation et de transfert de la matière organique. Ainsi :

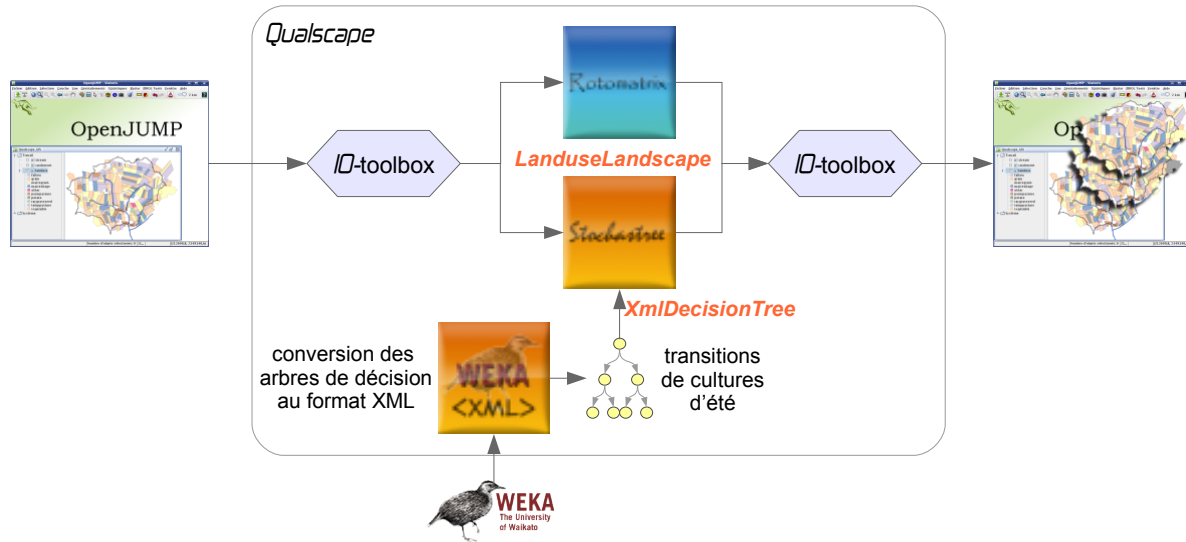
- au niveau des pratiques : les variations de l'efficacité des apports d'azote sont d'abord expliquées

par le système de culture ;

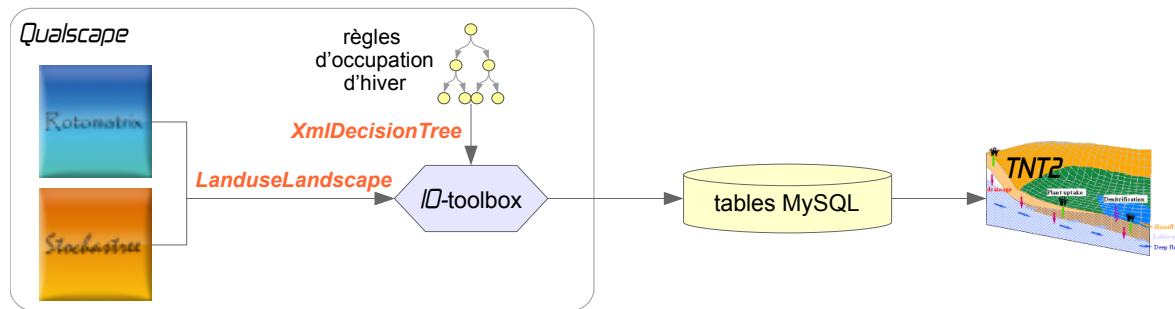
- au niveau des sols : les variations du bilan d'azote lixiviable, qui intègre les effets de la minéralisation et de la dénitrification sont d'abord expliquées par le climat puis par le système de cultures ;
- au niveau de l'exutoire du bassin versant : les variations de la concentration en nitrate à l'exutoire sont expliquées d'abord par le climat, puis par les sols et enfin par les pratiques.

De plus, les effets des interactions sol-climat sur cette concentration sont du même ordre qu'une réduction de 15% des apports totaux d'azote. Cela montre bien que la définition de pratiques agricoles alternatives devant satisfaire un objectif environnemental précis (comme le non-dépassement du seuil des 50 mg.l<sup>-1</sup> de nitrate dans les cours d'eau) ne peut se faire qu'en lien avec les conditions du milieu et que leur transposition directe d'un bassin versant à un autre peut ne pas atteindre l'effet escompté. Ces résultats sont néanmoins spécifiques au bassin versant choisi pour cette étude, où les pratiques intensives étaient caractérisées par une forte sur-fertilisation.

### 1. Simulation des transitions de cultures d'été, visualisation des résultats sous SIG



### 2. Simulation des transitions de cultures d'été, ajout des occupations d'hiver, export pour TNT2



### 3. Analyse des sorties « bassin versant » et parcellaires de TNT2, construction des bilans N et P

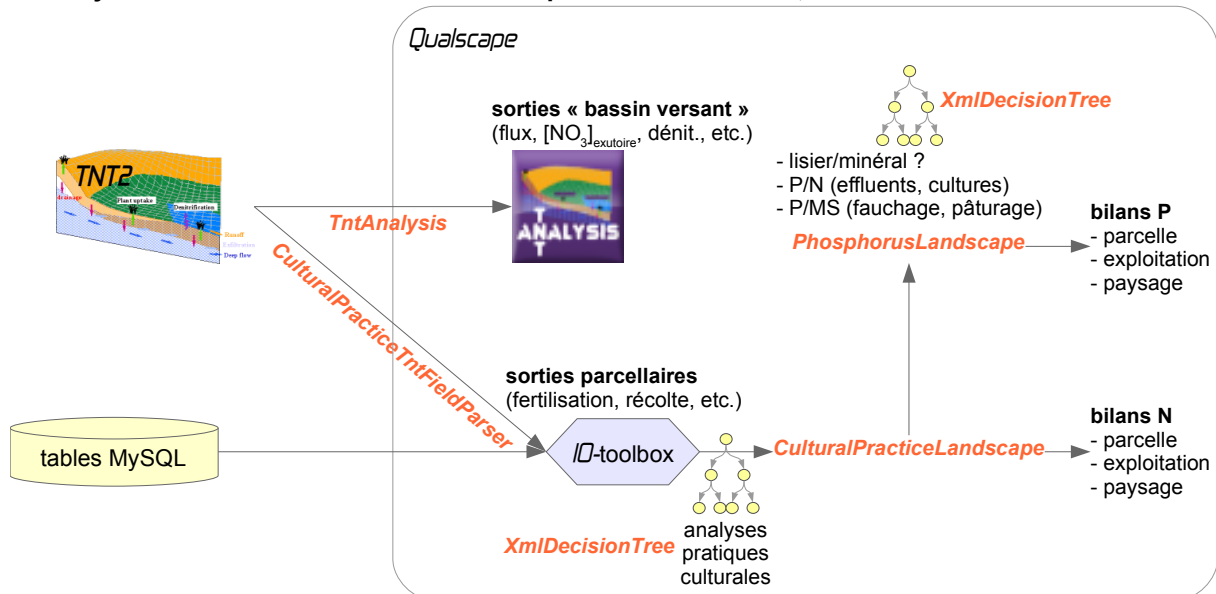


Fig. V.33 : Trois différents usages de Qualscape. La présence d'une icône carrée signale l'existence d'une interface graphique pour réaliser l'opération souhaitée. Les noms en orange désignent des classes majeures dans le fonctionnement de Qualscape. Les figurés arborescents indiquent l'utilisation locale d'arbres de décision. Les flèches symbolisent les flux de données généraux, mettant souvent en jeu la boîte à outils entrées-sorties de Qualscape (IO-toolbox).

## La plateforme Qualscape

Les modules de la plateforme *Qualscape* effectuent peu d'opérations calculatoires mais ont joué un rôle structurant dans la réalisation du projet à deux niveaux. Le premier a été de fournir une représentation orientée-objet et générique du paysage agricole (cf. chapitre II), qui a été complétée en fonction des besoins du projet. Ainsi, dans un contexte de simulation de l'occupation du sol, la classe *Field* contient une liste de *Landuse* dont le nombre est incrémenté à chaque nouvelle occupation du sol simulée. Cette structure est partagée par les modèles *Rotomatrix* et *Stochastree*. Par contre, dans un contexte d'analyse des pratiques culturales (les sorties parcellaires de TNT2 par exemple), la classe *Field* a été adaptée pour contenir une liste de *CulturalPractice*, classe générique déclinée en semis, fertilisation minérale, organique, récolte, etc. Une classe interprétant les sorties parcellaires de TNT2 a été développée afin de reconstruire l'architecture du paysage agricole et de permettre d'analyser ces sorties à l'échelle de la parcelle, de l'exploitation ou du paysage<sup>33</sup>. Cela nous amène au deuxième rôle structurant de *Qualscape* : celui d'interface entre la création de paysages virtuels, l'exportation de ces paysages vers TNT2, l'analyse des sorties de TNT2 et le couplage à un module phosphore. Ces fonctions d'interface sont présentées dans la figure V.33 qui récapitule différents usages possibles de *Qualscape* à ce stade du projet, ainsi que les modules correspondant mobilisés. La figure V.33 montre également que les arbres de décision ont joué des rôles cruciaux à différents stades du projet. D'un point de vue technique, une structure XML<sup>34</sup> a été définie pour les représenter, et une classe dédiée à leur manipulation a été conçue de façon à en permettre un usage polyvalent<sup>35</sup>. Cela nous a permis de condenser le code source des modules de *Qualscape* en conservant les opérations logiques qui leur sont propres dans le programme, et en externalisant sous forme d'arbres les structures décisionnelles susceptibles d'être modifiées par l'utilisateur : probabilités de transition de culture dans les arbres utilisés par *Stochastree*, règles d'allocation de l'occupation du sol hivernale, doses de phosphore apportées. L'utilisation d'arbres de décision comme structure logique externe à un programme nous paraît intéressante et porteuse dans l'évaluation environnementale des agrosystèmes puisque les démarches de fertilisation raisonnée combinent généralement des indicateurs de pratiques (a-t-on fertilisé l'an passé ? A-t-on fait une impasse ?) et d'état du système (la teneur en phosphore du sol est-elle supérieure ou inférieure à tel seuil ?) (Schvartz *et al.*, 2005).

---

33 il s'agit de la classe *CulturalPracticeTntFieldParser*

34 voir annexes III

35 il s'agit de la classe *XmlDecisionTree*, voir annexes III



## Points stratégiques

Cette partie récapitule soit des points non résolus dans notre démarche, soit des points nécessitant une certaine attention dans leur traitement.

### ***Stochasticité et déterminisme***

Nous avons mis en œuvre trois méthodes stochastiques dans la création des paysages virtuels ou dans la simulation de leur évolution :

- l'épaisseur des sols a été simulée à partir d'un variogramme expérimental extrait d'un jeu de données de 487 sondages. Pour chaque valeur simulée, l'algorithme employé combine le variogramme donné avec les valeurs observées dans le voisinage du point en question pour estimer une distribution de valeurs et en tirer une aléatoirement (Pebesma and Wesseling, 1998) ;
- les transitions de cultures sont simulées à partir de probabilités de transition (matrices ou arbres de décision) estimées à partir de 2409 successions de cultures observées sur le terrain ;
- les séries « climat sec » et « climat humide » ont été créées à partir d'une chronique de 13 années de données mesurées, en tirant aléatoirement parmi les six années les plus sèches et les plus humides respectivement.

Ainsi, pour chacune des combinaisons de facteurs *climat-milieu physique-système de culture*, une infinité de réalisations sont possibles. Or nous n'avons testé qu'une réalisation de chaque paysage virtuel. Des répétitions seraient nécessaires afin d'estimer l'influence des différentes sources de stochasticité sur l'évaluation environnementale d'un paysage virtuel.

Par ailleurs, les simulations de l'épaisseur des sols et des transitions de culture requièrent toutes les deux une première phase d'apprentissage dans la construction du variogramme expérimental et des arbres de décision. Or ces deux étapes sont réputées d'une part pour leur sensibilité à la qualité des données d'apprentissage, mais aussi par leur côté potentiellement trompeur dans le sens où même un apprentissage sur un mauvais jeu de données fournira de toute façon un variogramme ou un arbre de décision (Cressie, 1991 ; Quinlan, 1993). Une mauvaise qualité des données d'apprentissage alors que les simulations produites servent ensuite d'entrées à d'autres modèles entraîne des phénomènes de propagation d'erreurs ou d'incertitudes qui compliquent l'analyse et l'interprétation des résultats finaux (van Asselt et Rotmans, 2002).

### ***Validation de l'évaluation environnementale des paysages virtuels***

L'utilisation de la modélisation s'est imposée dans notre démarche pour créer les paysages virtuels et simuler de façon intégrée la diversité des processus d'évolution en jeu. Cependant, la validation des résultats de simulation est tout autant délicate car ils intègrent les effets de ces interactions, voire même utopique lorsqu'il n'y a pas de configuration réelle à confronter au paysage virtuel.

Cependant, afin de maintenir un certain niveau de vraisemblance dans notre démarche, nous avons élaboré les paysages virtuels à partir d'un paysage réel ayant déjà fait l'objet d'applications de TNT2 et compatible avec les hypothèses du modèle. Par ailleurs, le calage de TNT2 a été réalisé en comparant des données observées (débit et concentration) avec le paysage virtuel de référence (combinant le milieu physique, la chronique climatique et les pratiques agricoles représentant le paysage réel). Les autres paysages virtuels ont été construits en échantillonnant les chroniques climatiques et en réduisant l'intervalle de variation de l'épaisseur des sols. On peut ainsi considérer que les paysages virtuels alternatifs sont des sous-échantillons du paysage de référence, présentant donc une variabilité spatiale du milieu physique moindre et soumis à des séries climatiques dont l'intervalle de variation est compris dans la fourchette de valeurs rencontrées lors du calage de TNT2. Ainsi, ces paysages virtuels restent dans les limites fixées par les conditions de validité de TNT2. Cette démarche ne permet néanmoins que de créer une diversité restreinte de paysages virtuels.

La méthode mise en place pour évaluer et comparer les impacts environnementaux des agrosystèmes a nécessité des simulations de plusieurs décennies pour que la réponse hydrologique se stabilise et que la comparaison entre systèmes soit réalisée sur cette période stable. Cependant, Aspinall (2004) et Pocewicz et al. (2008) ont montré que les déterminants d'occupation du sol et des pratiques agricoles changeaient toutes les deux ou trois décennies, en fonction des changements socio-économiques locaux (échanges de parcelles, ventes d'exploitation) ou globaux (nouvelles réglementations, modification des équilibres économiques). Cela suggère que des simulations aussi longues à pratiques constantes ne sont pas réalistes d'un point de vue socio-économique, et que vraisemblablement peu de bassins-versants sont à l'état d'équilibre correspondant aux pressions environnementales auxquelles ils sont soumis. Cela nécessite donc une démarche d'analyse plus fine qui viserait à interpréter non pas un pseudo-état stationnaire, mais davantage les trajectoires d'évolution ; l'analyse de ces trajectoires supposerait néanmoins de bien appréhender les artefacts induits par les procédures d'initialisation des paysages virtuels et les variations causées par les aléas climatiques.

### ***Gestion de la matière organique***

La gestion de la matière organique est au cœur des problèmes environnementaux liés aux systèmes d'élevage intensifs. Se trouvant parfois en situation d'excédent, ces systèmes doivent trouver une façon de transférer ou de traiter leurs effluents. Notre approche a consisté à utiliser des itinéraires techniques type pour modéliser les pratiques agricoles, qui est un choix discutable à plusieurs niveaux. D'une part, il ne rend pas compte de la diversité et de la spécificité des exploitations appartenant à un même système de production. D'autre part, les itinéraires techniques types synchronisent artificiellement les opérations culturales indépendamment des conditions climatiques et de la disponibilité du matériel agricole. L'utilisation de modèle de fonctionnement d'exploitation d'élevage comme IFSM (Integrated Farm System Model) (Rotz et al., 2007) ou du projet « Mélodie »<sup>36</sup> (développé en partenariat avec l'INRA et l'IFIP) sont des alternatives à envisager pour simuler les processus décisionnels générant les pratiques culturales et intégrer les effets du climat et de la

---

36 Modélisation des Élevages en Langage Objet pour Déterminer leur Impact Environnemental

disponibilité du matériel sur la faisabilité des opérations culturelles.

Par ailleurs, la version de TNT2 utilisée dans le cadre de cette étude n'a pas permis de rendre compte de la diversité des caractéristiques des effluents qui se distinguaient par des C/N différents mais des dynamiques de décomposition identiques. Des développements récents ont été entrepris par l'équipe de développement de TNT2 afin de mieux rendre compte de cette diversité, mais n'ont pas pu être intégrés dans ce travail.

## **Perspectives – de l'évaluation environnementale à l'évaluation de la durabilité des agrosystèmes**

Ce projet représente une première tentative d'évaluation environnementale d'agrosystèmes par le couplage de paysages agricoles virtuels avec des modèles dynamiques et spatialisés. Deux ressources, l'eau et le sol, ont été suivies par des indicateurs axés sur deux critères, l'azote et le phosphore. L'intégration de nouveaux critères et de nouvelles ressources dans l'évaluation environnementale d'agrosystèmes, soulève de questions de recherche supplémentaires concernant la création des paysages virtuels, dans la méthodologie d'évaluation et dans les choix réalisés dans ce projet. Ces ouvertures nous permettront d'envisager comment les apports de ce projet pourraient contribuer à l'évaluation de la durabilité des agrosystèmes.

Une démarche d'évaluation intégrée des agrosystèmes suppose nécessairement l'intégration de ressources et d'indicateurs supplémentaires. Sans chercher l'exhaustivité, on peut évoquer en particulier : le transfert d'autres substances vers les sols et l'eau (pesticides, métaux lourds, carbone organique dissous), l'érosion des sols, les émissions de gaz toxiques à effet de serre, impacts sur la biodiversité, flux de pathogènes liés aux effluents, etc. Répondre à ces perspectives nécessite de considérer d'une façon plus systématique l'utilisation de résolutions spatiales et temporelles variables en fonction des objets et des processus à simuler.

Au niveau spatial, notre approche a combiné une représentation parcellaire (occupation du sol) et discrétisée par une grille régulière (flux d'eau). Cette combinaison convient également bien pour aborder les problématiques liées à l'application et aux transferts pesticides mais nécessite d'être complétée par une réflexion sur la connexion des parcelles en arbre d'exutoires. La représentation orientée-objet des parcelles que nous avons adoptée peut servir de base à une représentation plus fine pour fractionner les parcelles qui versent dans des parcelles différentes, notamment lorsque la parcelle-amont est sur une ligne de crête ou lorsque plusieurs parcelles aval reçoivent différents flux de la parcelle amont. La simulation des flux de surface – qu'il s'agisse d'eau, de matière ou d'êtres vivants – nécessite également la prise en compte d'éléments interstitiels comme les haies ou les fossés qui constituent des zones de transfert, d'échange ou de barrage privilégiées. Hormis l'intégration de ces nouveaux éléments (sous-parcelles, éléments interstitiels) dans notre modélisation du paysage, il nous faut également en développer les relations topologiques et sémantiques.

L'aspect temporel est fortement lié à l'échelle spatiale des processus simulés. Ainsi, notre approche a combiné un pas de temps annuel/saisonnier pour la simulation des transitions de cultures, et journalier pour la simulation des processus biophysiques et de transferts avec TNT2. Le pas de temps journalier est trop grossier pour simuler des processus plus gouvernés par des événements extrêmes comme l'érosion (sensible au pic d'intensité maximale de la pluie) ou fugaces comme la volatilisation d'azote lors du compostage en bâtiment ou de la restitution au pâturage. Si depuis l'avènement des SIG la manipulation d'objets d'échelle et de résolution spatiale différentes ne pose plus beaucoup de problèmes techniques, l'utilisation de *threads* – qui permet à différents objets de partager les mêmes ressources mais de façon asynchrone et

indépendante – pour simuler l'évolution des différents éléments du paysage est une voie à poursuivre, même si elle demeure encore délicate à mettre en œuvre. C'est d'ailleurs un point déterminant dans la gestion des flux d'informations si on envisage des couplages dynamiques entre *Qualscape* ou TNT2 avec d'autres modèles, en particulier avec des modèles décisionnels.

En effet, la réflexion sur la spatialisation des pratiques agricoles et leur réalisation dans le temps doit être approfondie. La discrétisation journalière du temps ne permet pas de savoir si un travail du sol, qui modifie l'état de surface du sol et enfouit la matière organique de surface, a lieu avant ou après l'événement pluvieux de la journée (il peut d'ailleurs y avoir plusieurs pluies dans une journée) et limite donc l'évaluation des risques d'érosion ou des flux associés au ruissellement (pesticides, effluents). Nous avons intégré la distance comme déterminant de l'occupation du sol et nous aurions pu l'utiliser également pour l'allocation des itinéraires techniques type afin de spatialiser les pratiques, car les agriculteurs fractionnent davantage les épandages sur les parcelles proches du corps d'exploitation et les regroupent sur les parcelles éloignées afin de réduire la charge de travail. Cependant, étant donné que l'utilisation d'itinéraires techniques type ne permet pas de tenir compte de l'effet du climat et de la portance des parcelles pour moduler les dates d'interventions culturales aux champs, il semble plus pertinent d'envisager l'utilisation de modèles décisionnels pour construire les interventions culturales utilisées par TNT2. C'est ce qui est envisagé dans le cadre d'un projet à venir dans l'UMR SAS où il est prévu de coupler TNT2 avec MELODIE. Les autres bénéfices envisageables avec ce type de couplage tiennent dans la modélisation explicite du fonctionnement de chaque exploitation agricole, qui nécessite en contrepartie de renseigner davantage des caractéristiques comme le cheptel de chaque exploitation ainsi que les objectifs de production et les besoins fourragers. Cela permettrait notamment de mieux estimer les quantités d'effluents épandues, l'état de minéralisation des effluents, et enfin, de mieux estimer les restitutions aux champs et les volatilisations associées. Au-delà d'une évaluation environnementale plus fine ou plus complète, ces développements nous paraissent également indispensables afin d'intégrer des indicateurs économiques liés aux intrants et aux productions pour évaluer la durabilité des agrosystèmes.

## **Bibliographie générale**

*« In all areas of science, the convergence and integration of information from different points of view, different disciplines, and different approaches are what lead to advances and breakthroughs in understanding. »*

Gene E. Likens

*« ± Un écrivain fait-il forcément un bon critique ?  
± Pas plus qu'un alcoolique ne fait forcément un bon barman. »*

Jim Bishop

# Bibliographie générale

- Aarts H.F.M., B. Habekotté, H. van Keulen. 2000. Phosphorus (P) management in the 'De Marke' dairy farming system. *Nutrient Cycling in Agroecosystems*, 56 (3):219-229.
- Agarwal C., G.M. Green, J.M. Grove, T.P. Evans, C.M. Schweik. 2002. A review and assessment of land-use change models: dynamics of space, time, and human choice. Gen. Tech. Rep. NE-297. Newton Square, PA: U.S. Department of Agriculture, Forest Service, Northeastern Research Station. 61 p.
- Agence canadienne d'évaluation environnementale. 2004. Évaluation environnementale stratégique. La directive du Cabinet sur l'évaluation environnementale des projets de politiques, de plans et de programmes. 18 pages.
- Alinat S. 2003. Paysages virtuels et projets territoriaux. 6<sup>es</sup> rencontres de Théo Quant. Février 2003:1-7.
- Anderson T.W. and L.A. Goodman. 1957. Statistical inference about Markov chains. *Annals of Mathematical Statistics*, 28 (1):89-110.
- Andrews S.S., D.L. Karlen, and C.A. Cambardella. 2004. The soil management assessment framework: A quantitative evaluation using case studies. *Soil Science Society of America Journal*, 68:1945-1962.
- Aspinall R. 2004. Modelling land use change with generalized linear models – a multi-model analysis of change between 1860 and 2000 in Gallatin Valley, Montana. *Journal of Environmental Management*, 72:91-103.
- Bachinger J. and P. Zander. 2007. ROTOR, a tool for generating and evaluating crop rotations for organic farming systems. *European Journal of Agronomy*, 26 (2):130-143.
- Baggs E.M., C.A. Watson, R.M. Rees. 2000. The fate of nitrogen from incorporated cover crop and green manure residues. *Nutrient Cycling in Agroecosystems*, 56 (2):153-163.
- Baize D., M.C. Girard. 1995. *Référentiel Pédologique*. INRA Éditions, 332 p.
- Baker W.L. 1989. A review of models of landscape change. *Landscape ecology*, 2 (2):111-133.
- Basset-Mens C., L. Anibar, P. Durand, H.M.G. van der Werf. 2006. Spatialised fate factors for nitrate in catchments: Modelling approach and implication for LCA results. *Science of The Total Environment*, 367 (1):367-382.
- Baudry J., C. Dupont, C. Thenail, V. Viaud. 2006. Aménagement du paysage et pratiques agricoles : quelles combinaisons dans la gestion des bassins versants ? In Mérot, P. *Qualité de l'eau en milieu rural : Savoirs et pratiques dans les bassins versants*. Inra Eds, Paris:123-135.
- Beaujouan V., P. Durand, L. Ruiz. 2001. Modelling the effect of the spatial distribution of agricultural practices on nitrogen fluxes in rural catchments. *Ecological Modelling*, 137:93-105.
- Beaujouan, V., P. Durand, L. Ruiz, P. Auousseau, G. Cotteret. 2002. A hydrological model dedicated to topography-based simulation of nitrogen transfer and transformation: rationale and application to the geomorphology-denitrification relationship. *Hydrological Processes*, 16:493-507.
- Berthe A.A., E. Carpenter, L. Codispoti, A.M. Izac, J. Lemoalle, F. Luizao, M. Scholes, P. Tréguer, B. Ward. 2005. Nutrient cycling. In *Ecosystems and human well-being: current state and trends*. Lavelle P., R. Dugdale, R. Scholes. Millennium Ecosystem Assessment. 948 pages:331-353.
- Beven K. 2001. *Rainfall-runoff modelling: the primer*. John Wiley & Sons, Chichester. 372 p.
- Beven K. 2002. Towards a coherent philosophy for modelling the environment. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 458 (2026):2465-2484.
- Beven K., Kirkby M. 1979. A physically based variable contributing area model of basin hydrology. *Hydrological Sciences Bulletin*, 24 (1):43-69.
- Bockstael, N., R. Costanza, I. Strand, W. Boynton, K. Bell, L. Wainger. 1995. Ecological economic modeling and valuation of ecosystems. *Ecological Economics*, 14:143-59.
- Bodet J.M., S. Hacala, C. Aubert, C. Texier. 2001. Fertiliser avec les engrais de ferme. Institut de l'Élevage, ITAVI, ITCF, ITP.
- Boiffin J. 1984. La dégradation structurale des couches superficielles du sol sous l'action des pluies. PhD thesis, Institut National Agronomique Paris-Grignon. 320 p.
- Bordenave P., F. Oehler, T. Bioteau, N. Turpin, P. Serrand, P. Saint-Cast, E. Le Saos. 2005. Distributed surface-groundwater coupled model applied to long term water management impacts at basin scale. *Ingénieries*, 44:13-36.



- Borrell V., I. Braud, G. Dedieu, A. Boone, F. Branger, Y. Brunet, I. Calmet, N. Carlier, A. Chanzy, Ph. Chibaudel, J.D. Creutin, H. Davi, A. Ern, F. Habets, F. Hecht, J. Jaffré, Ph. Lagacherie, J.C. Menaut, P. Mestayer, R. Moussa, J. Noilhan, J. Ogée, A. Olioso, L. Prévot, F. Rodriguez., M. Voltz. 2005. Modélisation du fonctionnement des surfaces continentales aux échelles locales à régionales SEVE : Sol Eau Végétation Energie, Actes du Colloque ECCO, 5-7 Décembre 2005, Toulouse:157-163.
- Brazier R.E., A.L. Heathwaite, S. Liu. 2005. Scaling issues relating to phosphorus transfer from land to water in agricultural catchments. *Journal of Hydrology*, 304:330-342.
- Breiman L., J.H. Friedman, R.A. Olshen, C.J. Stone. 1984. *Classification and regression trees*. Monterey, CA, Wadsworth.
- Brisson N., J. Wery, K. Boote. 2006. The two forms of crop models. *In* Wallach D., Makowski D., Jones J.W. Working with dynamic crop models: evaluation, analysis, parametrization, and applications. Elsevier Amsterdam:257-279.
- Brisson N., Mary B., Ripoche D., Jeuffroy M.H., Ruget F., Nicoullaud B., Gate P., Devienne-Barret F., Antonioletti R., Durr C., Richard G., Beaudoin N., Recous S., Tayot X., Plenet D., Cellier P., Machet J-M., Meynard J-M., Delécolle R. 1998. STICS: a generic model for the simulation of crops and their water and nitrogen balances. I. Theory and parameterization applied to wheat and corn. *Agronomie*, 18: 311-346.
- Brunet R. 1986. La carte modèle et les chorèmes. *Mappemonde*, 4:2-6.
- Burel F. and J. Baudry. 2003. *Landscape ecology: concepts, methods and applications*. Science Publishers, Inc. Enfield (NH), USA. 362 p.
- Burel F., J. Baudry, A. Butet, P. Clergeau, Y. Delettre, D. Le Coeur, F. Dubs, N. Morvan, G. Paillat, S. Petit, C. Thenail, E. Brunel and J.C. Lefeuvre, Comparative biodiversity along a gradient of agricultural landscapes, *Acta Oecologica*, 19 (1):47-60.
- Carpenter S.R., N.F. Caraco, D.L. Correll, R.W. Howarth, A.N. Sharpley, V.H. Smith. 1998. Nonpoint pollution of surface waters with phosphorus and nitrogen. *Ecological Applications*, 8 (3):559-568.
- Castellazzi M.S., G.A. Wood, P.J. Burgess, J. Morris, K.F. Conrad, J.N. Perry. 2007b. A systematic representation of crop rotations. *Agricultural Systems*, 97 (1-2):16-33.
- Castellazzi M.S., J.N. Perry, N. Colbach, H. Monod, K. Adamczyk, V. Viaud, K.F. Conrad. 2007a. New measures and tests of temporal and spatial pattern of crops in agricultural landscapes. *Agriculture, Ecosystems & Environment*, 118 (1-4):339-349.
- Cerdan O., V. Souchère, V. Lecomte, A. Couturier, Y. Le Bissonnais. 2002. Incorporating soil surface crusting processes in an expert-based runoff model: Sealing and Transfer by Runoff and Erosion related to Agricultural Management. *Catena*, 46 (2-3):189-205.
- Chabrier P., F. Garcia, R. Martin-Clouaire, G. Quesnel and H. Raynal. 2007. Toward a simulation modeling platform for studying cropping systems management: the Record project. Session Design, implementation, and application of software for model integration. *ModSim 07, New Zealand, Christchurch*, 10-13 December 2007. 8p.
- Chaplot V., C. Walter, P. Curmi. 2000. Improving soil hydromorphy prediction according to DEM resolution and available pedological data. *Geoderma*, 97 (3-4):405-422.
- Cheverry C. (ed.). 1998. *Agriculture intensive et qualité des eaux*. Science Update, Paris. 297 pages.
- Colin, F., C. Puech, G. de Marsily. 2000. Relations between triazine flux, catchment topography and distance between maize fields and the drainage network. *Journal of Hydrology*, 236 (3-4):139-152.
- Conover W.J. and R.L. Iman. 1981. Rank transformation as a bridge between parametric and nonparametric statistics: *The American Statistician*, 35 (3):124-129.
- Coppedge B.O., D.M. Engle, S.D. Fuhlendorf. 2007. Markov models of land cover dynamics in a southern Great Plains grassland region. *Landscape Ecology*, 22:1383-1393.
- Cressie N.A.C. 1991. *Statistics for spatial data*. John Wiley & Sons, New York. 900 p.
- Cressie N.A.C. 1991. *Statistics for spatial data*. John Wiley & Sons, New York. 900 p.
- D'Arcy B. and A. Frost. 2001. The role of best management practices in alleviating water quality problems associated with diffuse pollution. *The Science of the Total Environment*, 265:359-367.
- Deasy C., A.L. Heathwaite, R.E. Brazier. 2008. A field methodology for quantifying phosphorus transfer and delivery to streams in first order agricultural catchments. *Journal of Hydrology*, 350 (3-4):329-338.

- Deffontaines, J.P., C. Thenail, J. Baudry. 1995. Agricultural systems and landscape patterns: how can we build a relationship? *Landscape and Urban Planning*, 31 (1-3):3-10.
- DeFries R.S., J.C.W. Chan. 2000. Multiple criteria for evaluating machine learning algorithms for land cover classification from satellite data. *Remote Sensing of Environment*, 74 (3):503-515.
- Djordjic E., L. Bergstrom, C. Grant. 2005. Phosphorus management in balanced agricultural systems, *Soil Use and Management* 21, 94-101
- Dogliotti S., W.A.H. Rossing, M. K. van Ittersum. 2003. ROTAT, a tool for systematically generating crop rotations. *European Journal of Agronomy*, 19 (2):239-250.
- Dokuchaev V.V. 1883. Russian Chernozem. *In Selected Works of V.V. Dokuchaev*, Vol. 1, p. 14–419. Moscow, 1948. Israel Program for Scientific Translations Ltd. (for USDA-NSF), S. Monson, Jersalem, 1967. (Translated from Russian into English by N. Kaner).
- Doncaster, C.P. and A.J.H. Davey. 2007. *Analysis of variance and covariance. How to choose and construct models for the life sciences.* Cambridge University Press, Cambridge. 302 pages.
- Dray S. and A.B. Dufour. 2007. The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software*, 22(4):1-20.
- Dubus I.G., C.D. Brown, S. Beulke. 2003. Sources of uncertainty in pesticide fate modelling. *The Science of The Total Environment*, 317 (1-3):53-72.
- Durand P., C. Hénault, J. Bidois, F. Trolard. 1998. La dénitrification en zone humide de fonds de vallée. *In Cheverry C. (ed.) Agriculture intensive et qualité des eaux.* Science Update, Paris:223-231.
- Durand P., F. Ferchaud, J. Salmon-Monviola, F. Goetschel, C. Martin. 2006. Étude sur l'évolution des paramètres nitrates dans les eaux brutes des bassins versants Bretagne Eau Pure et des autres bassins versants bretons. Rapport final, volets 3 et 4. INRA, Bretagne Eau Pure.
- Ellis E.C. and N. Ramankutty. 2008. Putting people in the map: anthropogenic biomes of the world. *Frontiers in Ecology and the Environment* 6, doi:10.1890/070062.
- European Commission. 1991. Directive of the Council of 12 December 1991 concerning the protection of waters against pollution caused by nitrates from agricultural sources (91/676/EEC). European Commission, Brussels.
- European Commission. 1999. Directions towards sustainable agriculture. Communication to the Council, the European Parliament, the Economic and Social Committee and the Committee of the Regions, COM (1999) 22 Final, January 1, 1999. [http://europa.eu.int/comm/agriculture/envir/9922/9922\\_en.pdf](http://europa.eu.int/comm/agriculture/envir/9922/9922_en.pdf).
- FAO. 1991. The Den Bosch declaration and agenda for action on sustainable agriculture and rural development. Report on the FAO/Netherlands conference on agriculture and the environment, 'S-Hertogenbosch, The Netherlands, 15-19 April, 1991, FAO, Rome.
- Fardeau, J.C., and A.P. Conesa. 1994. Le phosphore. *In Bonneau, M. & Souchier, B. Pédologie 2. Constituants et propriétés du sol:* 557-568.
- Finnveden G. and J. Potting. 1999. Eutrophication as an impact category – state of the art and research needs, *International Journal of Life Cycle Assessment* 4 (6):311-314.
- Forman R.T.T. 1995. Some general principles of landscape and regional ecology. *Landscape Ecology*, 10:133-142.
- Forman R.T.T. and M. Gordon. 1981. Patches and structural components for a landscape ecology. *BioScience*, 31:733-740.
- Fournier F. and C. Cheverry. 1992. Les échelles spatiales d'étude du rôle du sol dans l'environnement. *In Auger P., J. Baudry, F. Fournier. Hiérarchies et échelles en écologie:*21-41.
- Franzluebbers A.J., H.H. Schomberg, D.M. Endale. 2007. Surface-soil responses to paraplowing of long-term no-tillage cropland in the Southern Piedmont USA. *Soil and Tillage Research*, 96 (1-2):303-315.
- Friedl M.A. and Brodley, C.E. 1997. Decision tree classification of land cover from remotely sensed data. *Remote Sensing of Environment*, 61 (3):399-409.
- Friedl M.A., D.K. McIver, J.C.F. Hodges, X.Y. Zhang, D. Muchoney, A.H. Strahler, C.E. Woodcock, S. Gopal, A. Schneider, A. Cooper, A. Baccini, F. Gao and C. Schaaf. 2002. Global land cover mapping from MODIS: algorithms and early results. *Remote Sensing of Environment*, 83 (1-2):287-302.
- Gabrielle B. 2006. L'évaluation environnementale des agrosystèmes : une approche intégrée pour gérer les risques agri-

- environnementaux. Mémoire d'Habilitation à Diriger les Recherches, Université Paris VI. 58 p.
- Gabrielle B., B. Mary, R. Roche, P. Smith, G. Gosse. 2002. Simulation of carbon and nitrogen dynamics in arable soils: a comparison of approaches. *European Journal of Agronomy*, 18:107-120.
- Gaucherel C., D. Fleury, D. Auclair, P. Dreyfus. 2006b. Neutral models for patchy landscapes. *Ecological Modelling*, 197 (1-2):159-170.
- Gaucherel C., F. Burel, J. Baudry. 2007. Multiscale and surface pattern analysis of the effect of landscape pattern on carabid beetles distribution. *Ecological Indicators*, 7 (3):598-609.
- Gaucherel C., J. Baudry, D. Auclair, F. Houllier. 2004. État de l'art de la modélisation du paysage. Rapport INRA, SAD EFPA, AMAP. 31 p.
- Gaucherel C., N. Giboire, V. Viaud, T. Houet, J. Baudry, F. Burel. 2006a. A domain-specific language for patchy landscape modelling: the Brittany agricultural mosaic as a case study. *Ecological Modelling*, 194:233-243.
- Gladwin C.H. 1989. *Ethnographic decision tree modeling*. Sage, Newbury Park.
- Goovaerts P. 1997. *Geostatistics for natural resources evaluation*. Oxford University Press, New York.
- Grizzetti B., F. Bouraoui, G. De Marsily. 2005. Modelling nitrogen pressure in river basins: a comparison between a statistical approach and the physically-based SWAT model. *Physics and Chemistry of the Earth*, 30 (8-10):508-517.
- Groggel D.J. and J.H. Skillings. 1986. Distribution-free tests for main effects in multifactor designs. *The American Statistician*, 40:99-102.
- Gustafson E.J. and T.R. Crow. 1998. Simulating spatial and temporal context of forest management using hypothetical landscapes. *Environmental Management*, 22 (5):777-787.
- Haas G., F. Wetterich, U. Köpke. 2001. Comparing intensive, extensified and organic grassland farming in southern Germany by process life cycle assessment. *Agriculture, Ecosystems & Environment*, 83 (1-2):43-53.
- Halberg N., H.M.G. van der Werf, C. Basset-Mens, R. Dalgaard, I.J.M. Boer. 2005. Environmental assessment tools for the evaluation and improvement of European livestock production systems. *Livestock Production Science*, 96 (1):33-50.
- Hanink D.M. and R.G. Cromley. 1998. Land-use allocation in the absence of complete market values. *Journal of Regional Science*, 38 (3):465-480.
- Hart M.R., B.F. Quin, M.L. Nguyen. 2004. Phosphorus runoff from agricultural land and direct fertilizer effects: a review. *Journal of Environmental Quality*, 33 (6):1954-1972.
- Haynes R.J. 1986. *Mineral nitrogen in the plant-soil system*. Ed. Kozlowsky T.T. Madison (USA). 483 p.
- Hazen G.B. 2002. Stochastic tree and the StoTree modeling environment: models and software for medical decision analysis. *Journal of Medical Systems*, 26 (5):399-413.
- Heathwaite L., A.N. Sharpley, M. Bechmann. 2003. The conceptual basis for a decision support framework to assess the risk of phosphorus loss at the field scale across Europe. *Soil Science and Plant Nutrition*, 166 (4):447-458.
- Heidenberger K. 1996. Dynamic project selection and funding under risk: a decision tree based MILP approach. *European Journal of Operational Research*, 95:284-298.
- Heijungs R., M. Goedkoop, J. Struijs, S. Effting, M. Sevenster, G. Huppes. 2003. Towards a life cycle impact assessment method which comprises category indicators at the midpoint and the endpoint level. Report of the first project phase: design of the new method. 74 p. Disponible sur : [http://www.leidenuniv.nl/interfac/cml/ssp/publications/recipe\\_phase1.pdf](http://www.leidenuniv.nl/interfac/cml/ssp/publications/recipe_phase1.pdf)
- Helsel D.R. and R.M. Hirsch. 2002. *Statistical methods in water resources*. USGS, Techniques of Water-Resources Investigations Book 4, Chapter A3.
- Hénault C. and J.C. Germon. 2000. NEMIS, a predictive model of denitrification on the field scale. *European Journal of Soil Science*, 51:257-270.
- Horton R.E. 1933. The role of infiltration in the hydrological cycle. *Transactions of the American Geophysical Union*, 14:446-460.
- Houdart M., Bonin M., Saudubray F. 2004. Typology of spatial farms management: the Capot watershed in Martinique. *Les Cahiers de la Multifonctionnalité*, 5:57-69.
- Houet T. and L. Hubert-Moy, 2006. Modelling and projecting land-use and land-cover changes with cellular automaton

- in considering landscape trajectories: an improvement for simulation of plausible future states. *EARSeL eProceedings*, 5 (1):63-76.
- Houghton P.D., Charman P.E.V. 1986. Glossary of terms, used in soil conservation. Soil Conservation Service of New South Wales.
- Huang Q.H., Y.L. Cai, J. Peng. 2007. Modeling the spatial pattern of farmland using GIS and multiple logistic regression: a case study of Maotiao River Basin, Guizhou Province, China. *Environmental Modeling and Assessment*, 12 (1):55-61.
- Hunter M.D. 2002. Landscape structure, habitat fragmentation, and the ecology of insects. *Agricultural and Forest Entomology*, 4 (3):159-166.
- Hutchings N.J., J.E. Olesen, B.M. Petersen, J. Berntsen. 2007. Modelling spatial heterogeneity in grazed grassland and its effects on nitrogen cycling and greenhouse gas emissions. *Agriculture, Ecosystems & Environment*, 121:153-163.
- IUSS Working Group WRB. 2007. World Reference Base for Soil Resources 2006, first update 2007. World Soil Resources Reports No 103. FAO, Rome.
- Johnson M. E. 1987. *Multivariate Statistical Simulation*. Wiley, New York, 230 p.
- Jordan M.I., Z. Ghahramani, L.K. Saul. 1997. Hidden Markov decision trees. *In* M. C. Mozer, M.I. Jordan, T. Petsche (Eds.), *Advances in Neural Information Processing Systems (NIPS) 10*, Cambridge MA: MIT Press.
- Kanchikerimath M. and D. Singh. 2001. Soil organic matter and biological properties after 26 years of maize-wheat-cowpea cropping as affected by manure and fertilization in a Cambisol in semiarid region of India. *Agriculture, Ecosystems & Environment*, 86 (2):155-162.
- Karlen D.L., E.G. Hurley, S.S. Andrews, C.A. Cambardella, D.W. Meek, M.D. Duffy, A.P. Mallarino. 2006. Crop rotation effects on soil quality at three northern corn/soybean belt locations. *Agronomy Journal*, 98:484-495.
- Karpinets T.V., D.J. Greenwood, J.T. Ammons. 2004. Predictive mechanistic model of soil phosphorus dynamics with readily available inputs. *Soil Science Society of America Journal*, 68 (2):644-653.
- Kelly R.H., W.J. Parton, G.J. Crocker, P.R. Graced, J. Klír, M. Körschens, P.R. Poulton, D.D. Richter. 1997. Simulating trends in soil organic carbon in long-term experiments using the century model. *Geoderma*, 81(1-2):75-90.
- Kingery W.L., C.W. Wood, J.C. Williams. 1996. Tillage and amendment effects on soil carbon and nitrogen mineralization and phosphorus release. *Soil and Tillage Research*, 37 (4):239-250.
- Kirkby M.J. 1978. *Hillslope hydrology*. John Wiley & Sons, Chichester.
- Ladet S., M. Deconchat, C. Monteil, J.P. Lacombe, G. Balent. 2005. Les chaînes de Markov spatialisées comme outil de simulation. *Géomatique, Cassini*, 04:159-173.
- Lagacherie P. 1992. Formalisation des lois de distribution des sols pour automatiser la cartographie pédologique a partir d'un secteur pris comme référence. PhD thesis, Université Montpellier II, France.
- Lal R. 2004. Soil carbon sequestration to mitigate climate change. *Geoderma*, 123 (1-2):1-22.
- Largouët C. and M.O. Cordier. 2001 Improving the landcover classification using domain knowledge. *AI Communication*, 1:35-43.
- Lawrence R., A. Bunn, S. Powell, M. Zambon. 2004. Classification of remotely sensed imagery using stochastic gradient boosting as a refinement of classification tree analysis. *Remote Sensing of Environment*, 90:331-336.
- Le Ber F., A. Napoli, J. Metzger, S. Lardon. 2003. Modeling and comparing farm maps using graphs and case-based reasoning. *Journal of Universal Computer Science*, 9 (9):1073-1095.
- Le Ber F., M. Benoit, C. Schott, J.F. Mari, C. Mignolet. 2006. Studying crop sequences with CarrotAge, a HMM-based data mining software. *Ecological Modelling*, 191 (1):170-185.
- Le Bissonnais Y. and D. Arrouays. 1997. Aggregate stability and assessment of soil crustability and erodibility: II. Application to humic loamy soils with various organic carbon contents. *European Journal of Soil Science*, 48 (1):39-48.
- Leenhardt D., D. Wallach, P. Le Moigne, M. Guérif, A. Bruand, M.A. Casterad. 2006. Using crop models for multiple fields. *In* Wallach D., Makowski D., Jones J.W. Working with dynamic crop models: evaluation, analysis, parametrization, and applications. Elsevier Amsterdam:209-244.
- Lemercier B., C. Walter, C. Schwartz, N. Saby, D. Arrouays, S. Follain. 2006. Organic carbon and extractable

- phosphorus contents monitoring in cultivated soils of three French regions. *Étude et Gestion des Sols*, 73 (3):165-179.
- Lewis D.R. and M.B. McGehan. 2002. A review of field scale phosphorus dynamic models. *Biosystems Engineering*, 82 (4):359-380.
- Li X. and A.G.O. Yeh. 2002. Neural-network-based cellular automata for simulating multiple land use changes using GIS. *International Journal of Geographical Information Science*, 16 (4):323-343.
- Liang H., H. Zhang, Y. Yan. 2006. Decision Trees for Probability Estimation: An Empirical Study. NRC 48783. The 18th IEEE International Conference on Tools with Artificial Intelligence (ICTAI06). Washington D.C., USA. November 13-15, 2006. 10 pages.
- Liu X.J., A.R. Mosier, A.D. Halvorson, C.A. Reule, F.S. Zhang. 2007. Dinitrogen and N<sub>2</sub>O emissions in arable soils: Effect of tillage, N source and soil moisture. *Soil Biology and Biochemistry*, 39 (9):2362-2370.
- Logofet D.O. and E.V. Lesnaya. 2000. The mathematics of Markov models: what Markov chains can really predict in forest successions? *Ecological Modelling*, 126:285-298.
- Lynn H., C.L. Mohler, S.D. DeGloria, C.E. McCulloch. 1995. Error assessment in decision-tree models applied to vegetation analysis. *Landscape Ecology*, 10 (6):323-335.
- Mary B. and J. Rémy. 1979. Essai d'appréciation de la capacité de minéralisation de l'azote des sols de grande culture. I. Signification des cinétiques de minéralisation de la matière organique humifiée. *Annales Agronomiques*, 30:513-527.
- Mary B., S. Recous, S. Darwis, D Robin. 1996. Interactions between decomposition of plant residues and nitrogen cycling in soil. *Plant Soil*, 181-71-82.
- McBratney A.B., M.L. Mendonça Santos, B. Minasny. 2003. On digital soil mapping. *Geoderma*, 117 (1-2):3-52.
- McGuire K.J. and J.J. McDonnell. 2006. A review and evaluation of catchment transit time modeling. *Journal of Hydrology*, 330 (3-4):543-563.
- Mérot P., C. Gascuel-Oudou, C. Walter, X. Zhang, J. Molénat. 1995. Bocage landscape and surface water pathways. *Revue des sciences de l'eau*, 12 (1):23-44.
- Mignolet C., C. Schott, M. Benoît. 2004. Spatial dynamics of agricultural practices on a basin territory: a retrospective study to implement models simulating nitrate flow. The case of the Seine basin. *Agronomy for Sustainable Development*, 24:219-236.
- Mignolet C., C. Schott, M. Benoit. 2007. Spatial dynamics of farming practices in the Seine basin: Methods for agronomic approaches on a regional scale. *Science of The Total Environment*, 375 (1-3):13-32.
- Molénat J. and C. Gascuel-Oudou. 2002. Modelling flow and nitrate transport in groundwater for the prediction of water travel times and of consequences of land use evolution on water quality. *Hydrological Processes*, 16 (2):479-492.
- Molénat J., C. Gascuel-Oudou, L. Ruiz, G. Gruau. 2008. Role of water table dynamics on stream nitrate export and concentration in agricultural headwater catchment (France). *Journal of Hydrology*, 348 (3-4):363-378.
- Mosier A. and C. Kroeze. 2000. Potential impact on the global atmospheric N<sub>2</sub>O budget of the increased nitrogen input required to meet future global food demands. *Chemosphere, Global Change Science*, 23 (4):465-473.
- Moussa R., M. Voltz, P. Andrieux. 2002. Effects of the spatial organization of agricultural management on the hydrological behaviour of a farmed catchment during flood events. *Hydrological Processes*, 16 (2):393-412.
- Murthy S.K. 1998. Automatic construction of decision trees from data: a multi-disciplinary survey. *Data Mining and Knowledge Discovery*, 2:345-389.
- Nash, J.E. and J.V. Sutcliffe. 1970. River flow forecasting through conceptual models. *Journal of Hydrology*, 10:282-290.
- Nicolardot B., B. Mari, S. Houot, S. Recous. 1996. La dynamique de l'azote dans les sols cultivés. *In* Maîtrise de l'azote dans les agrosystèmes. 19-20 novembre 1996, Reims (France):87-103.
- Öborn I., A.C. Edwards, E. Witter, O. Oenema, K. Ivarsson, P.J.A. Withers, S.I. Nilsson, A. Richert Stinzing. 2003. Element balances as a tool for sustainable nutrient management: a critical appraisal of their merits and limitations within an agronomic and environmental context. *European Journal of Agronomy*, 20 (1-2):211-225.
- Oehler F., P. Bordenave, P. Durand. 2007. Variations of denitrification in a farming catchment area. *Agriculture*,

- Ecosystems & Environment, 120 (2-4):313-324.
- Oenema O., H. Kros, W. de Vries. 2003. Approaches and uncertainties in nutrient budgets: implications for nutrient management and environmental policies. *European Journal of Agronomy*, 20 (1-2):3-16.
- Paillat J.M., P. Robin, M. Hassouna, P. Leterme. 2005. Predicting ammonia and carbon dioxide emissions from carbon and nitrogen biodegradability during animal waste composting. *Atmospheric Environment*, 39 (36):6833-6842.
- Pal M. and P.M. Mather. 2003. An assessment of the effectiveness of decision tree methods for land cover classification. *Remote Sensing of Environment*, 86:554-565.
- Pauwels H., P. Lachassagne, P. Bordenave, J.C. Foucher, A. Martelat. 2001. Temporal variability of nitrate concentration in a schist aquifer and transfer to surface waters. *Applied Geochemistry*, 16:583-596.
- Payraudeau S. and H.M.G. van der Werf. 2005. Environmental impact assessment for a farming region: a review of methods. *Agriculture, Ecosystems & Environment*, 107 (1):1-19.
- Payraudeau S., H.M.G. van Der Werf, F. Vertès. 2006. Evaluation of an operational method for the estimation of emissions of nitrogen compounds for a group of farms. *International Journal of Agricultural Resources, Governance and Ecology*, 5 (2-3):224-246.
- Payraudeau S., H.M.G. van der Werf. 2005. Environmental impact assessment for a farming region: a review of methods. *Agriculture, Ecosystems & Environment*, 107 (1):1-19.
- Pebesma E.J. and Wesseling C.G. 1998. Gstat: a program for geostatistical modelling, prediction and simulation. *Computers & Geosciences*, 24 (1):17-31.
- Pebesma, E.J. 2004. Multivariable geostatistics in S: the gstat package. *Computers & Geosciences*, 30 (7):683-691.
- Pennock D.J., A. Veldkamp. 2006. Advances in landscape-scale soil research. *Geoderma*, 133 (1-2):1-5.
- Perlich C., F. Provost, J. S. Simonoff. 2003. Tree induction vs. logistic regression: a learning-curve analysis. *Journal of Machine Learning Research*, 4 (2):211-255.
- Peterjohn W.T. and W.P. Carson. 1984. Nutrient dynamics in an agricultural watershed: observations on the role of a riparian forest. *Ecology*, 65:1466-1475.
- Pocewicz A., M. Nielsen-Pincus, C. Goldberg, M. Johnson, P. Morgan, J. Force, L. Waits, L. Vierling. 2008. Predicting land use change: comparison of models based on landowner surveys and historical land cover trends. *Landscape Ecology*, 23 (2):195-210.
- Portejoie S., J. Martinez, G. Landmann. 2002. L'ammoniac d'origine agricole : impacts sur la santé humaine et animale et sur le milieu naturel. *INRA Productions Animales*, 15(3):151-160.
- Potting J. and K. Blok. 1994. Spatial aspects of life-cycle assessment. *In* Udo de Haes H.A., A.A. Jensen, W. Klöpffer, L.G. Lindfors (Ed). *Integrating impact assessment into LCA*, Society of Environmental Toxicology and Chemistry (SETAC) – Europe, Belgium, Brussels.
- Préfecture d'Ille et Vilaine. 2005. Arrêté du 23 novembre 2005 relatif au troisième programme d'action à mettre en oeuvre en vue de la protection des eaux contre la pollution par les nitrates d'origine agricole. 58 pages.
- Provost F. and P. Domingos. 2003. Tree induction for probability-based ranking. *Machine Learning*, 52 (3):199-215.
- Quinlan J. R. 1981, Induction of decision trees. *Machine Learning*, 1:81-106.
- Quinlan J.R. 1990. Probabilistic decision trees. *In* *Machine Learning: An Artificial Intelligence Approach - Volume III*, Y. Kodratoff and R.S. Michalski, Eds. Morgan Kaufmann, San Mateo, CA, 140-152.
- Quinlan J.R. 1993. *C4.5: Programs for Machine Learning*. San Francisco: Morgan Kaufmann.
- R Development Core Team. 2007. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.r-project.org>.
- Rocchini D. and C. Ricotta. 2007. Are landscapes as crisp as we may think? *Ecological Modelling*, 204:535-539.
- Rotz C.A., M.S. Corson, C.U. Coiner. 2007. The integrated farm system model reference manual, version 2.1. Pasture systems and watershed management research unit, Agricultural research service, United States Department of Agriculture. 141 p.
- Rounsevell M.D.A., J.E. Annetts, E. Audsley, T. Mayr, I. Reginster. 2003. Modelling the spatial distribution of agricultural land use at the regional scale. *Agriculture, Ecosystems & Environment*, 95 (2-3):465-479.
- Ruiz L., P. Arousseau, J. Baudry, V. Beaujouan, P. Cellier, P. Curmi, P. Durand, C. Gascuel-Oudou, P. Leterme, J.L.

- Peyraud, C. Thenail, C. Walter. 2006. Conception de bassins-versants virtuels : vers un outil pour l'étude de l'influence de l'organisation spatiale de l'activité agricole et du milieu physique sur les flux d'azote dans les bassins-versants. *In* Monestiez, P., S. Lardon, B. Seguin. *Organisation spatiale des activités agricoles et processus environnementaux*. Update Sciences & technologies, INRA, Paris:337-354.
- Ruiz L., S. Abiven, P. Durand, C. Martin, F. Vertès, V. Beaujouan. 2002. Effect on nitrate concentration in stream water of agricultural practices in small catchments in Brittany: I. Annual nitrogen budgets, *Hydrology and Earth System Sciences*, 6 (3):497-506.
- Russel E.W. 1973. *Soil condition and plant growth*. 10<sup>th</sup> ed. Longmans, London.
- Salmon-Monviola J., Durand P., Ferchaud F., Sorel L. 2008. Land cover and cropping system generation: a modelling tool to construct scenarios as input to a distributed nitrate transfer model. Submitted to *Environmental Modelling & Software*.
- Schröder J.J., H.F.M. Aarts, H.F.M. ten Berge, H. van Keulen, J.J. Neeteson. 2003. An evaluation of whole-farm nitrogen balances and related indices for efficient nitrogen use. *European Journal of Agronomy*, 20 (1-2):33-44.
- Schvartz C., J.C. Muller, J. Decroux. 2005. *Guide de la fertilisation raisonnée*. France Agricole (ed.).
- Seaman J.W., S.C. Walls, S.E. Wise, R.G Jaeger. 1994. Caveat emptor: rank transform methods and interaction. *Trends in Ecology & Evolution*, 9:261-263.
- Shannon C.E. 1948. A mathematical theory of communication. *Bell System Technical Journal*, 27:379-423, 623-656.
- Shapiro S.S. and Wilks M.B. 1965. An analysis of variance test for normality. *Biometrika*, 52 (3-4):591-611.
- Sharpley A.N., B. Foy, P.J.A. Withers. 2000. Practical and innovative measures for the control of agricultural phosphorus losses to water: an overview. *Journal of Environmental Quality*, 29:1-9.
- Snedecor G.W. and Cochran W.G. 1989. *Statistical Methods*, Eighth Edition, Iowa State University Press.
- Sonneveld M.P.W., J.M. Schoorl, A. Veldkamp. 2006. Mapping hydrological pathways of phosphorus transfer in apparently homogeneous landscapes using a high-resolution DEM. *Geoderma*, 133 (1-2):32-42.
- Sørensen R. and J. Seibert. 2007. Effects of DEM resolution on the calculation of topographical indices: TWI and its components. *Journal of Hydrology*, 347 (1-2):79-89.
- Strahler A.N. 1952. Dynamic basis of geomorphology. *Geological Society of America Bulletin*, 63:923-938.
- Sumfleth K. and R. Duttman. 2008. Prediction of soil property distribution in paddy soil landscapes using terrain data and satellite information as indicators. *Ecological Indicators*, 8 (5):485-501.
- Tarim S.A., S. Manandhar, T. Walsh. 2006. Stochastic constraint programming: a scenario-based approach. *Constraints*, 11 (1):53-80.
- Thenail C. and J. Baudry. 1994. Méthodes d'étude des relations entre activités agricoles et paysages. *In* Sébillotte M. (Ed.), *Recherches-Systèmes en Agriculture et Développement Rural*; 13e Symposium International. CIRAD Editions, Montpellier:316-321.
- Thenail C. and J. Baudry. 2004. Variation of farm spatial land use pattern according to the structure of the hedgerow network (bocage) landscape: a case study in northeast Brittany. *Agriculture, Ecosystems & Environment*, 101 (1):53-72.
- Thornton P.K. and P.G. Jones. 1998. A conceptual approach to dynamic agricultural land-use modelling. *Agricultural Systems*, 7:505-521.
- Tortrat F., P. Aourousseau, H. Squiquidant, C. Gascuel-Oudou, M.O. Cordier. 2004. Modèle Numérique d'Altitude (MNA) et spatialisation des transferts de surface : utilisation de structures d'arbres reliant les exutoires de parcelles et leurs surfaces contributives. *Bulletin SFPT*, 172, 128-136.
- Trepos R. 2008. Apprentissage symbolique à partir de données issues de simulation pour l'aide à la décision. Gestion d'un bassin versant pour une meilleure qualité de l'eau. PhD thesis. Université de Rennes 1. 139 p.
- Trepos R., A. Salleb, M.O. Cordier, V. Masson, C. Gascuel-Oudou. 2005. A distance approach for action recommendation. *In* *Machine Learning: ECML 2005 - Proceedings*. Gama J., R. Camacho, P. Brazdil, A. Jorge, L. Torgo (Eds). 16th European Conference on Machine Learning. 3-7 Oct. 2005, Porto, PT. Vol. 3720:425-436.
- Turner M.G., R.H. Gardner, R.V. O'Neil. 2001. *Landscape ecology in theory and practice: Pattern and process*. Springer-Verlag, New York, New York, USA.
- US GCRP. 2003. *Our changing planet, FY 2003*. A report by the climate change science program and the subcommittee

on global change research. 124 p.

- Usher M.B. 1979. Markovian approaches to ecological succession. *Journal of Animal Ecology*, 48:413-426.
- Usher M.B. 1992. Statistical models of succession. *In* Glenn-Lewin DC, Peet RK, Veblen TT (eds). *Plant succession: theory and prediction*. Chapman & Hall, London:215-248.
- Vadas P.A., R.D. Harmel, P. J. A. Kleinman. 2007. Transformations of soil and manure phosphorus after surface application of manure to field plots. *Nutrient Cycling in Agroecosystems*, 77:83-99.
- Vadas P.A., W.J. Gburek, A.N. Sharpley, P.J.A. Kleinman, P.A. Moore, M.L. Cabrera, R.D. Harmel. 2007. A model for phosphorus transformation and runoff loss for surface-applied manures. *Journal of Environmental Quality*, 36 (2):324-332.
- Van Asselt M.B.A. and J. Rotmans. 2002. Uncertainty in Integrated Assessment Modelling. *Climatic Change*, 54 (1):75-105.
- Van der Werf H.M.G. and J. Petit. 2002. Evaluation of the environmental impact of agriculture at the farm level: a comparison and analysis of 12 indicator-based methods. *Agriculture, Ecosystems & Environment*, 93 (1-3):131-145.
- Van Ittersum M.K., F. Ewert, T. Heckeles, J. Wery, J. Alkan Olsson, E. Andersen, I. Bezlepkina, F. Brouwer, M. Donatelli, G. Flichman, L. Olsson, A.E. Rizzoli, T. van der Wal, J. E. Wien, J. Wolf. 2008. Integrated assessment of agricultural systems - A component-based framework for the European Union (SEAMLESS). *Agricultural Systems*, 96 (1-3):150-165.
- Veldkamp A. and L.O. Fresco. 1996. CLUE: a conceptual model to study the Conversion of Land Use and its Effects. *Ecological Modelling*, 85:253-270.
- Verburg P.H. 2006. Simulating feedbacks in land use and land cover change models. *Landscape Ecology*, 21:1171-1183.
- Verburg P.H., P.P. Schot, M. J. Dijst, A. Veldkamp. 2004. Land use change modelling: current practice and research priorities. *Geojournal*, 61:309-324.
- Viaud V. 2004. Organisation spatiale des paysages bocagers et flux d'eau et de nutriments. Approche empirique et modélisations. PhD thesis, École Nationale Supérieure Agronomique de Rennes.
- Viaud V., P. Mérot, J. Baudry. 2004. Hydrochemical buffer assessment in agricultural landscapes: from local to catchment scale. *Environmental Management*, 34 (4):559-573.
- Vought L.B.M., J. Dahl, C.L. Pedersen, J.O. Lacoursiere. 1994. Nutrient retention in riparian ecotones. *Ambio*, 23:342-348.
- Wackernagel M., L. Onisto, P. Bello, A. Callejas Linares, I. Susana López Falfán, J. Méndez García, A. Isabel Suárez Guerrero, M. Guadalupe Suárez Guerrero. 1999. National natural capital accounting with the ecological footprint concept. *Ecological Economics*, 29 (3):375-390.
- Wallach D. 2006. The two forms of crop models. *In* Wallach D., Makowski D., Jones J.W. Working with dynamic crop models. Evaluation, analysis, parametrization, and applications. Elsevier Amsterdam:3-10.
- Walter C. and P. Curmi. 1998. Les sols du bassin versant du Coët-Dan : organisation, variabilité spatiale et cartographie. *In* Cheverry C. (ed.). *Agriculture intensive et qualité des eaux*. Science Update: Paris:85-105.
- Walter C., D. King, P. Lagacherie, J.M. Robbez-Masson. 2005. L'analyse spatiale des sols : description, modélisation et représentation de la variabilité spatiale des sols. *In* Girard M.C., C. Walter, J.C. Remy, J. Berthelin, J.L. Morel (Eds). *Sols et environnement. Cours et Étude de Cas*. DUNOD. Sciences Sup., chapitre 8.
- Walter C., R. Rossel, A. Viscarra, A.B. McBratney. 2003. Spatio-temporal simulation of the field-scale evolution of organic carbon over the landscape. *Soil Science Society of America Journal*, 67:1477-1486.
- Weaver K. and A.H. Perera. 2003. Modelling land cover transitions: A solution to the problem of spatial dependence in data. *Landscape Ecology*, 19 (3):273-289.
- West T.O. and W.M. Post. 2002. Soil organic carbon sequestration rates by tillage and crop rotation: a global data analysis. *Soil Science Society of America Journal*, 66 (6):1930-1946.
- Wielemaker W.G., S. de Bruin, G.F. Epema, A. Veldkamp. 2001. Significance and application of the multi-hierarchical landsystem in soil mapping. *Catena*, 43 (1):15-34.
- Williams J.R., C.A. Jones, P.T. Dyke. 1984. Modeling approach to determining the relationship between erosion and soil productivity. *Transactions of the American Society of Agricultural Engineers*, 27 (1):129-144.
- Witten I.H. and E. Frank. 2005. *Data Mining: Practical machine learning tools and techniques*, 2<sup>nd</sup> Edition. Morgan



Kaufmann, San Francisco.

- Wu F. and C.J. Webster. 1998. Simulation of land development through the integration of cellular automata and multicriteria evaluation. *Environment and Planning B: Planning and Design*, 25 (1):103-126.
- Yuan Y. and M.J. Shaw. 1995. Induction of fuzzy decision trees. *Fuzzy Sets and Systems*, 69 (2):125-139.
- Zahm F., P. Girardin, C. Mouchet, P. Viaux, L. Vilain. 2005. From the assessment of the sustainability of farms with IDEA method to the characterization of the European agriculture with IDERICA method. *Colloque international sur les Indicateurs Territoriaux du Développement Durable*, Université Paul Cézanne (Aix en Provence), 1-2 décembre 2005:1-17.
- Zhang C. and W. Li. 2008. A comparative study of nonlinear Markov chain models for conditional simulation of multinominal classes from regular classes. *Stochastic Environmental Research and Risk Assessment*, 22 (2):217-230.
- Zhang C. and W. Li. 2008. A comparative study of nonlinear Markov chain models for conditional simulation of multinominal classes from regular classes. *Stochastic Environmental Research and Risk Assessment*, 22 (2):217-230.

## **Annexes**

### **Annexes I : Article soumis à Journal of Environmental Quality (en révision)**

Application d'une modélisation par paysage virtuel pour évaluer des stratégies d'échantillonnage d'agroécosystèmes (résumé français)

Application of virtual landscape modeling to assess agroecosystem monitoring strategies (article en anglais)

### **Annexes II : Matrices de transition et arbres de décision stochastiques relatifs aux résultats des chapitres III et IV**

Matrices de transition entre les cultures pour les différents types d'exploitation du bassin versant de Naizin

Arbres de décision utilisés par Stochastree pour les différents types d'exploitation du bassin versant de Naizin

Arbre de décision utilisé pour simuler l'interculture hivernale

### **Annexes III : Format et manipulation des arbres de décision**

Définition du format XML utilisé

Production et visualisation des arbres

Méthode principale de manipulation des arbres



## **Annexes I**

Article soumis à Journal of Environmental Quality (en  
révision)



# **Annexes I : Application d'une modélisation par paysage virtuel pour évaluer des stratégies d'échantillonnage d'agroécosystèmes<sup>37</sup>**

## **Résumé français**

Le suivi dans le temps de l'évolution des sols d'un agroécosystème peut être rendue généralement difficile par la forte variabilité spatiale des sols, l'importance des variations temporelles de court ou de long termes et l'étendue des zones d'étude considérées. Néanmoins, des contraintes réglementaires se développent pour assurer un suivi de la qualité des sols dans l'optique d'une préservation de cette ressource ou pour vérifier l'efficacité de programmes environnementaux.

Il apparaît de ce fait utile de pouvoir tester l'efficacité prévisionnelle de réseaux de surveillance des sols avant leur mise en place effective et en tenant de la variabilité du milieu considéré et de la dynamique supposée des propriétés considérées. L'objectif de ce travail est de développer une méthode fondée sur la construction de paysages virtuels réalistes, permettant de tester la précision de plusieurs stratégies d'échantillonnage des sols au cours du temps et leur aptitude à déceler les tendances d'évolution.

En raison des enjeux environnementaux associés, la teneur en phosphore des sols a été retenue comme variable d'intérêt. Une distribution spatiale réaliste des teneurs en phosphore a été générée par une approche de simulation stochastique, sur un paysage de 15 x 15 km dont le parcellaire et la topographie sont décrits avec une résolution de 50 m (Walter et al., 2003) : cette distribution respecte des structures spatiales connues (de courte, moyenne et longue portées) des variations des teneurs en phosphore dans les sols. L'évolution de l'occupation du sol, les apports d'effluents organiques et les bilans de phosphore ont été modélisés sur une période de 50 ans en supposant un changement important dans les pratiques de fertilisation au bout de 20 ans de simulation. Ce modèle d'évolution spatio-temporelle de la teneur en phosphore des sols a constitué un paysage virtuel, sur lequel nous avons appliqué différentes stratégies d'échantillonnage (grille systématique, stratification par l'occupation du sol initiale, stratification par l'occupation du sol courante, par transects) selon des densités et des pas de temps croissants.

Confrontées à l'évolution parfaitement connue du paysage virtuel, les stratégies d'échantillonnage non stratifiées sont apparues plus biaisées que les stratégies fondées sur une stratification par l'occupation du sol, qui est le déterminant majeur de l'initialisation et de l'évolution des teneurs en P. Les stratégies fondées sur un réseau permanent de sites de mesure sont apparues plus efficaces pour détecter des tendances. In fine, la stratification par l'occupation du sol initiale est apparue dans le cas de figure étudié comme la meilleure option, car sa pondération tenait le mieux compte de l'historique des pratiques agricoles. Ce travail de

---

37 Article soumis à Journal of Environmental Quality et en révision

comparaison met également en évidence l'importance potentielle des biais et des risques de détection erronées des tendances d'évolution, quand des stratégies non pertinentes ou fondées sur un échantillonnage trop faible dans l'espace et dans le temps, sont mises en œuvre.

L'approche ainsi développée, fondée sur l'évaluation prévisionnelle des performances d'une stratégie d'échantillonnage sur un paysage virtuel présentant des structures spatiales et temporelles analogues au paysage réel, apparaît de ce fait comme un outil innovant d'aide à la conception de programmes de surveillance efficaces.

# Application of virtual landscape modeling to assess agroecosystem monitoring strategies

Luc Sorel<sup>a</sup>, Christian Walter<sup>a,\*</sup>, Thomas Chauvet<sup>a</sup>, Alex McBratney<sup>b</sup>,

<sup>a</sup> INRA, Agrocampus Rennes, UMR 1069, Soil, Agro- and hydro- Systems, F-35000 Rennes

<sup>b</sup> ACPA, The University of Sydney, NSW 2006, Australia

Abbreviations: BDAT, French Soil Analysis Database; DEM, digital elevation model; P, phosphorus; NST, normal score transformation; RF, random function; DE, dispersion envelope; IL, initial landuse; CL, current landuse; RG, regular grid; Tr, transect; CI, confidence interval; DE, dispersion envelope.

\* Corresponding author: christian.walter@agrocampus-rennes.fr INRA-Agrocampus Rennes, UMR SAS, 65 rue de Saint Briec, CS 84215, 35042 Rennes Cedex, France (phone: 0(+33) 2-23-48-59-42 / fax: 0(+33) 2-23-48-54-30)



## **Abstract**

The monitoring of soil in an agroecosystem appears difficult to carry out due to high spatial variability, short and long term temporal variations and large areal extent. Additionally, increasing regulatory constraints require the implementation of survey programs to assess environmental policies efficiency. To assist the effective design of monitoring networks in a context where only little data on the surveyed area is available, we developed a method based on realistic virtual landscapes to test the accuracy of several sampling strategies and their ability to detect trends.

Considering soil and water quality issues, soil phosphorus content was chosen as an important variable to be studied. A realistic spatial distribution of soil phosphorus was generated on a 15 km by 15 km area, using known spatial dependence structures. Landuse evolution, manure spreading and soil phosphorus budget were modelled over 50 years with a shift of fertilization practices after the 20<sup>th</sup> year. This model of spatio-temporal evolution of soil phosphorus content yielded a virtual landscape on which we tested different design-based sampling strategies.

Strategies based on self-weighting designs were significantly more biased than strategies based on designs stratified by landuse, which was a driving factor for the initialization and the evolution of the landscape. Stationary designs being more efficient in detecting trends, stratification by initial landuse stood as the best option because its weighting accounted best for agricultural practices history. The methodological approach on virtual landscapes and the strategy comparisons are transferable to environmental stakeholders willing to efficiently design survey programs.

## Introduction

The increasing development of human activities is raising the profile of environmental issues and human societies are more and more concerned about both primary resource out-takes and waste in-takes. In the context of an agroecosystem, the balance of fluxes between the soil nutrient export by crops and fertilizer applications is crucial for a sustainable development. Therefore many countries have set legal regulations and environmental survey programs to monitor and study the temporal trends of diverse environmental resources: the Environmental Monitoring Assessment Program (EMAP) surveys ecological resources all over the US (Lesser et al., 1997), the Soil Action Plan (SAP), and the Soil Quality Network (RMQS) were set-up to map and monitor soils in England (DEFRA, 2004) and in France (King et al., 2005), respectively.

However, environmental survey programs are technically burdensome to implement. The main difficulty is due to the highly variable nature of the environmental characteristics: they are spatially distributed both horizontally and vertically, particularly for soil attributes. Secondly, many characteristics are likely to vary in time. Therefore, to get an accurate idea of the state and the evolution of environment resources, a survey program requires a large number of spatially distributed samples, implying a high cost due to sampling set up, data analysis, and technician and hardware transportation. This financial burden is often a reason to reduce the financial resources allocated to survey programs (Lesser et al., 1997; King et al., 2005). Therefore, to avoid an overestimation of the means required by a survey program, environmental stakeholders need tools to facilitate their design.

The compromise between the efficiency and the cost of a survey program can be optimized if, beforehand, survey program designers have a better idea of the spatial variability and the dynamics of the surveyed variables. Two characteristics of survey programs can be optimized with respect to the spatial variability of the surveyed variables: the sampling density (number of samples by unit area) and the sampling design (the method used to place a sampling site on the study area). Moreover, the sampling frequency can be optimized according to the evolution process dynamics of the surveyed variables. Sampling strategies can either be design-based (simple random sampling, stratified random sampling, grid sampling) or model-based (sampling based on the geostatistical estimation of the spatial variability of the studied variable) (De Gruijter et al., 2006; Brus et al., 1997). Simulation is a relevant way of designing and testing sampling strategies before their real implementation. Firstly, simulation is cost-efficient because it only requires computing resources, and secondly, it allows the implementation and the assessment of different policy scenarios (Carriquiry et al., 1998). However, the environment to be sampled must be simulated too, either from existing data or from simulated landscapes. Several approaches to optimize a sampling strategy by simulation exist:

- optimization of the accuracy/cost ratio: in this case, the objective is to use the smallest number of sampling locations to estimate with an acceptable accuracy the mean value or the distribution of a variable over the studied region. Such optimization approaches often follow model-based strategies

(McBratney et al., 1983), subsampling of existing environmental databases (Fernández et al., 2005), or using spatial correlations of auxiliary variables (Simbahan et al., 2006);

- trend detection ability: optimization research on this issue is based either on field experiments or on simulated distribution. Field-based studies compare the instantaneous spatial variability of one or several characteristics measured on-site with its variability through time (Anderson A.B., 2002; Conant et al., 2003; Saby et al., 2004), whereas simulation-based approaches test different sampling strategies over a simulated distribution derived from existing sample databases (Lesser et al., 1997) or from modelled spatial structures like variograms (Papritz et al., 1995).

Results derived from existing environmental databases are site specific and do not try to formalize the spatial structures of the sampled variables and their evolution processes. Therefore, such sampling assessments are hardly transferable to other geomorphological, climatic, and anthropogenic conditions. Moreover, the trend detection approach is hardly studied mainly because of the short span of time covered by the dataset. On the other hand, most of the spatial distributions used in design optimizations are too simplistic in the way they model the spatial structure of their studied variable. Firstly, the spatial distribution is often modelled roughly with only one variogram, thus ignoring sources of variability which have either a finer or a coarser resolution. Secondly, the evolutionary processes barely consider both biophysical processes and anthropogenic actions. Therefore, our objectives were (i) to stochastically create a realistic virtual landscape by considering the influence of topography on field pattern, the effect of the landuse history on soil P content, and the combination of multiscale soil P content semivariograms (intra-field and regional scale), (ii) to simulate the evolution of this landscape and associated agricultural practices at an annual timestep over a 50 years period, (iii) to implement different survey programs, and (iv) to assess monitoring efficiency in revealing the impacts of agricultural practices on the environment.

We chose soil phosphorus (P) content as the target variable because it is a vital fertilizer for crops, and it represents a major issue for soil and water quality. P fertilization has improved considerably crop yields (Fardeau and Conesa, 1994) in agricultural practices since the late XIX<sup>th</sup> century, especially in areas where soils were lacking in P (Boulaine, 1996). However, agricultural P fertilization is estimated to be responsible for the soil P content increase (Némery et al., 2001) and is mainly responsible for P loadings in surface waters in many countries nowadays (McDowell et al., 2001; Heathwaite et al., 2003; Dougherty et al., 2004).

## Material and methods

### *Construction of the initial state of the virtual landscape*

#### Stochastic field pattern and landuse

The field pattern and the initial landuse of the virtual landscape were designed following Walter et al. (2003). The study area is a 15 km by 15 km region of Brittany (north-west France) for which we had a Digital Elevation Model (DEM) spatially lumped in a 300 by 300 pixel regular grid, each square cell being 50 m by 50 m. The field pattern was generated through the use of a variance quadtree algorithm (Minasny et al., 2007) applied on the DEM.

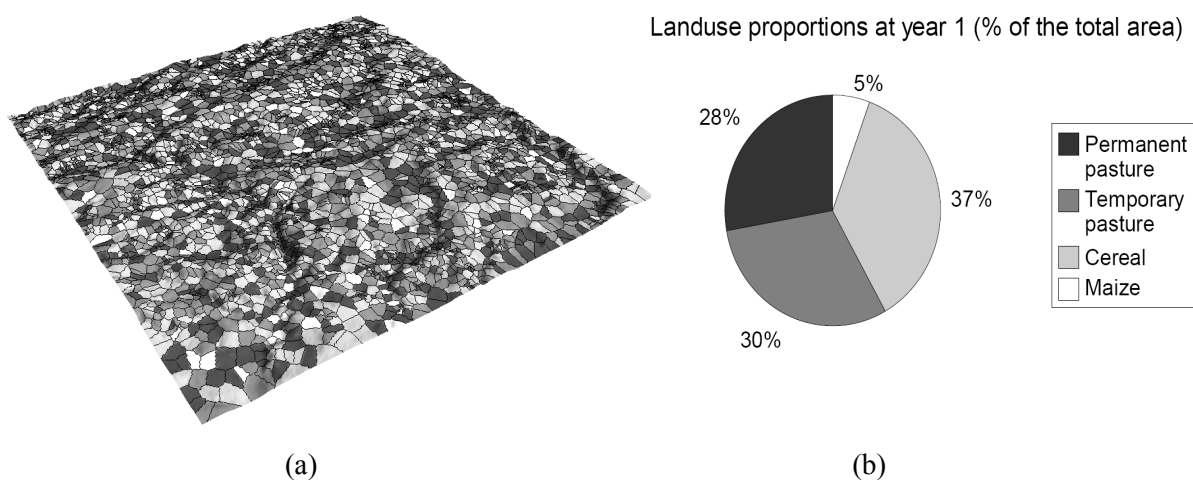


Fig. A.I.34: Initial landuse of the virtual landscape: (a) field pattern of the virtual landscape draped on the DEM. The landuse represents the situation in the early 60s (year 1 of the simulation). The landscape extent is 15km by 15km; (b) landuse proportion at year 1 (% of the total area).

In order to produce smaller fields in steep areas, the algorithm increased the field density proportionally to the variance of altitude, which accounts for realistic features observed in agricultural basins. 4153 fields were created, with areas ranging from <1 to 12 ha (Fig. A.I.34.a). To comply with the field pattern derived from the DEM, other nominal variables like landuse were rasterized at the spatial resolution of the DEM (50 m by 50 m). Spatial and numeric variables described hereafter related to soil P content were discretized at the same resolution, which is fine enough to account for soil P content intra-field variability.

Four landuses were considered in this study: permanent pastures, temporary pastures, cereals and maize. The initial landuse of each field was set randomly with two constraints: the proportions should be similar to the landuse in Brittany in the 1960's, and pastures were preferentially located in the valleys (Walter et al., 2003). The initial proportions were roughly similar for permanent pastures, temporary pasture and cereals (between 30 and 40% of the landscape area) and 5% for maize (Fig. A.I.34.b).

## Stochastic and deterministic components of the soil P content initial distribution

Among the various P compounds present in the soil, we focused on the plant available P reservoir, on which most of fertilization recommendations and soil databases rely (Aurousseau, 2001; Schwartz et al., 2005). The initial soil P content landscape was built using three different scales of spatial variability. Two experimental variograms were used to stochastically derive short- and long-range variability components, and a deterministic approach was used to model a medium-range component which accounted for the effect of field history.

A short-range variogram was derived from a nested survey of a 6 ha cultivated field located in Essonnes (Central France) and sampled on 70 sites in 1996 (ITCF, unpublished data). We derived from this data set a short-range spatial structure (0 to 120 m), which accounted for intra-field variability (equation (1)). A long-range variogram was derived from the French Soil Analysis Database (BDAT) which brought together more than 147,000 soil analyses over a large part of western France, including the study area (Walter et al., 1997) (equation (2)). To preserve the anonymity of the farmers, the spatial resolution of the database was downscaled to the township of the soil analyses. We randomly reallocated the analyses in a radius of 1 km around their township to derive a regional variogram, thus reliable for a distance between 2000 to 25,000 m. The spatial structures derived from the soil databases were fitted by least square approximation to the P experimental semi-variances using Splus (Kaluzny et al., 1998), leading to the following spherical and exponential variograms:

$$\gamma_{short\ range}(h) = \begin{cases} 0 + 1050 \left( \frac{3}{2} \left( \frac{h}{115} \right) - \frac{1}{2} \left( \frac{h}{115} \right)^3 \right), & \text{if } 0 \leq h \leq 115 \\ 1050, & \text{if } h > 115 \end{cases} \quad (1)$$

$$\gamma_{long\ range}(h) = 5300 + 4200 \left( 1 - \exp\left( \frac{-h}{11\ 500} \right) \right) \quad (2)$$

Geostatistical simulations of P at these scales were conducted over the landscape using equations (1) and (2) to derive respective random function (RF) models that would be representative of the study area (Fig. A.I.35.a). The generation of each corresponding short- and long-range spatial component had to be separated into two stages due to the large number of values to simulate (90,000 pixels) with the available statistical tools: the *rfsim* function of SPLUS (Kaluzny et al., 1998) was used to model a gross distribution (300 by 300 m pixels) with the Cholesky decomposition method (Cressie, 1991), and then the resolution was increased (50 by 50 m pixels) by sequential Gaussian simulations (Goovaerts, 1997) using KRAIG, a software designed by the Australian Centre for Precision Agriculture.

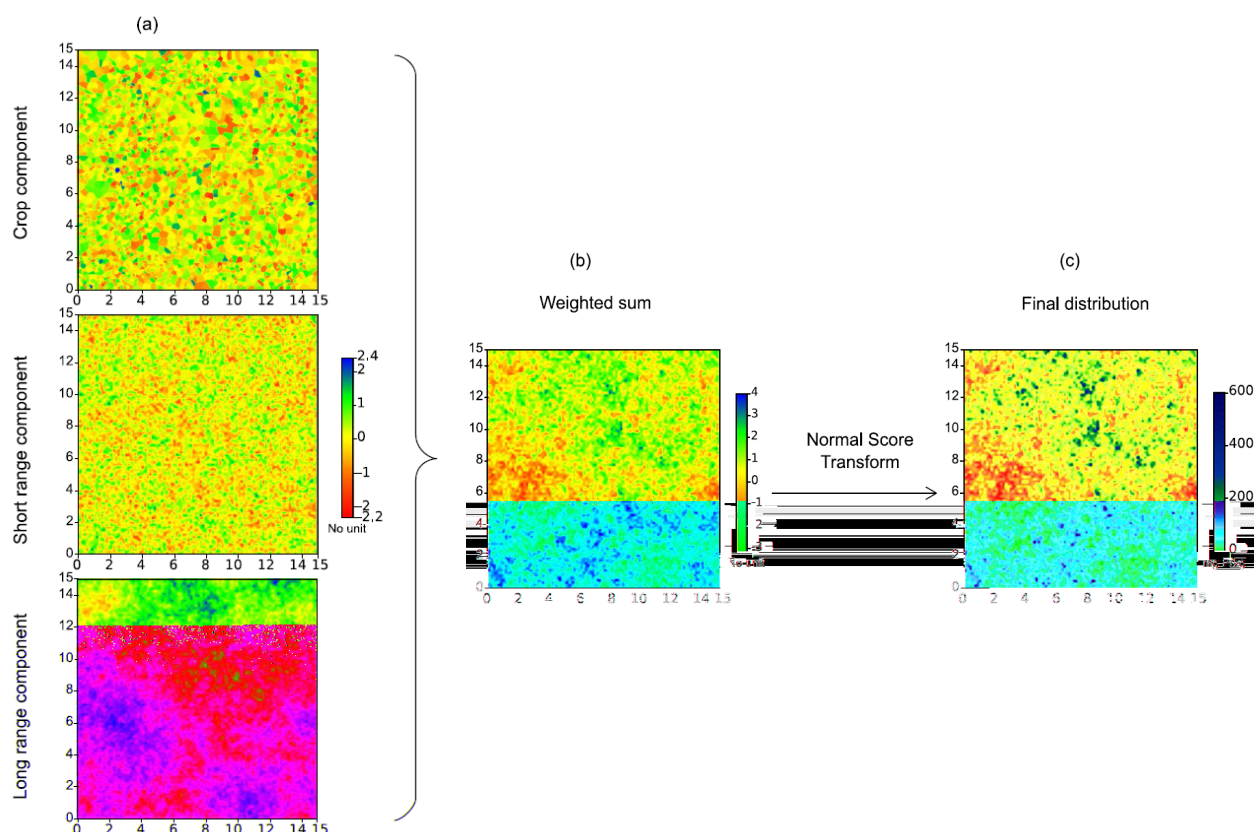


Fig. A.I.35: Construction of the initial soil P content distribution: (a) deterministic generation of the crop component and stochastic generation of the short and long range components; (b) weighted sum of the spatial distributions; (c) final soil P distribution corrected by the Normal Score Transformation.

We created a medium-range crop component which accounted for inter-field variability. Assuming that (i) pastures were less fertilized than other crops and that (ii) pastures are usually set for a long period of time, each field of the initial landuse was attributed a “cultural” P value according to four normal distributions corresponding to the four landuses considered here. Parameters of the normal distribution (Table A.I.10) were derived from regional references (Schvartz et al., 2005; Lemerrier et al., 2006). The field pattern generated by the selected realization is shown in Fig. A.I.35.a. Before combination with the other spatial components, this realization was transformed into a standard normal distribution using a Normal Score Transform (NST) (Sprent, 1989; Goovaerts, 1997).

Crop	Soil P content (mg P/kg )	
	Mean	Standard deviation
Permanent pasture	60	6
Temporary pasture	110	11
Cereal	110	11
Maize	180	18

Table A.I.10: Normal distribution parameters considered for the cultural component of the initial soil P content distribution.

To allow a permissible description of the overall spatial variability of soil P content in the study area, the two RF that describe the short- and long-range variability were combined with the medium-range spatial field using a linear model of regionalization (Goovaerts, 1997; Walter et al., 2003). The combined spatial field  $S(x)$  was derived from a linear combination of the two independent RFs and the deterministic spatial field, so that:

$$S(x) = \sum_{i=1}^3 a_i Y_i(x) \quad (3)$$

where  $x$  is a point of the area,  $Y_i(x)$  is one of the three spatial field ( $i=1$  is the medium-range crop component,  $i=2$  is the short-range component,  $i=3$  is long-range component). Each RF and spatial field  $Y_i(x)$  has zero mean and semivariogram  $\gamma_i(h)$ . Therefore, the combined permissible semivariogram was described as a linear combination of each semivariogram model  $\gamma_i(h)$ , so that:

$$\gamma(h) = \sum_{i=1}^3 b_i \gamma_i(h) \quad (4)$$

where the positive coefficient  $b_i$  is the variance contribution of the respective standardized semivariogram models  $\gamma_i(h)$ , and  $b_i = (a_i)^2$ . To estimate the  $b_i$ , the short- and long-range RF components and the deterministic medium-range component were weighted according to the variance contribution of each spatial structure. The nugget of the long-range variogram (5300 (mg P/kg)<sup>2</sup>) was decomposed into variances of the two other spatial structures. The short-range experimental variogram had a sill of 1050 (mg P/kg)<sup>2</sup>. Ten realizations of the medium-range component were created and the median sill of the variograms computed on these realizations was 1300 (mg P/kg)<sup>2</sup>. Therefore, the variance contributions of equation (4) were estimated as follows:

$$b_{long\ range} = \frac{sill_{long\ range}}{nugget_{long\ range} + sill_{long\ range}} = \frac{4200}{4200 + 5300} = 0.44 \quad (5)$$

$$b_{medium\ range} = (1 - b_{long\ range}) \frac{VAR_{medium\ range}}{VAR_{short\ range} + VAR_{medium\ range}} = 0.56 \frac{1300}{1050 + 1300} = 0.31 \quad (6)$$

$$b_{short\ range} = (1 - b_{long\ range}) \frac{VAR_{short\ range}}{VAR_{short\ range} + VAR_{medium\ range}} = 0.56 \frac{1050}{1050 + 1300} = 0.25 \quad (7)$$

The linear model of regionalisation produced a normal distribution of soil P content (Fig. A.I.35.b). A backward NST was applied to adapt the histogram of the simulated values to the frequency distribution derived from the BDAT database (Walter et al., 1997). The final soil P content distribution is shown on Fig. A.I.35.c.

## **Virtual landscape evolution**

Three interacting processes influenced the evolution of the soil P landscape: annual landuse evolution, annual fertilization practices and a soil P budget accounting for fertilization out-takes, crop in-takes, and soil P immobilization.

### **Stochastic landuse evolution**

We used a transition probability matrix to model the landuse evolution year after year as a probabilistic Markov chain (Cressie, 1991). We therefore assumed that landuse allocation was a random function only depending on the previous landuse. The probabilities can be represented as follows:

$$\begin{bmatrix} P_{Pp \rightarrow Pp} & P_{Pp \rightarrow Tp} & P_{Pp \rightarrow C} & P_{Pp \rightarrow M} \\ P_{Tp \rightarrow Pp} & P_{Tp \rightarrow Tp} & P_{Tp \rightarrow C} & P_{Tp \rightarrow M} \\ P_{C \rightarrow Pp} & P_{C \rightarrow Tp} & P_{C \rightarrow C} & P_{C \rightarrow M} \\ P_{M \rightarrow Pp} & P_{M \rightarrow Tp} & P_{M \rightarrow C} & P_{M \rightarrow M} \end{bmatrix}$$

where  $p_{i \rightarrow j}$  is the probability of changing from a landuse  $i$  to a landuse  $j$  between year  $y$  and  $y+1$ . Pp stands for permanent pasture, Tp for temporary pasture, C for cereal, and M for maize. The probabilities were designed to model the global trend of landuse evolution observed in this region between the 1960's and today: a diminution of pastures, an augmentation of maize fields and a rather constant number of temporary pastures and cereal fields (Table A.I.11). This flexible approach allows the design of different landuse evolution scenarios by solely changing the transition probabilities (Walter et al., 2003).



$$\begin{bmatrix} 0.95 & 0.035 & 0.01 & 0.005 \\ 0.01 & 0.65 & 0.24 & 0.1 \\ 0 & 0.2 & 0.8 & 0 \\ 0 & 0.09 & 0.01 & 0.9 \end{bmatrix}$$

Table A.I.11: Landuse transition probability matrix

<i>Manure origin</i>	<i>P unit for 1 N unit</i>
cattle	0.3
pig	0.48
poultry	0.36

(a)

<i>Crop</i>	<i>Exportations</i>		<i>Expected yields</i>	
			<i>Mean</i>	<i>Standard deviation</i>
Temporary pasture	3.1 kg P/t dry matter		3.5 t dry matter/ha	0.35 t dry matter/ha
Permanent pasture	3.1 kg P/t dry matter	X	7 t dry matter/ha	0.7 t dry matter/ha
Cereal	5 kg P/t		7.5 t/ha	0.75 t/ha
Maize	2.4 kg P/t dry matter		14 t dry matter/ha	1.4 t dry matter/ha

(b)

	<i>P thresholds (mg P/kg )</i>	
	<i>P<sub>reinforced</sub></i>	<i>P<sub>skip</sub></i>
<i>Demanding crops</i>		
consecutive cereals		
maize	100	300
pasture		
<i>Non-demanding crops</i>		
first cereal		250

(c)

Table A.I.12: Numerical values used in the P-based fertilization method: (a) P/N ratios of different manures, (b) P exportation coefficient by types of crop and expected yields by types of crop. The standard deviation column was used only to stochastically compute the actual crop P exportation, (c) Preinforced and Pskip values for a few crops in Brittany (mg P/kg ).

## Fertilization practices

We considered fertilization practices based on organic manures, which are characterized by a given balance of Nitrogen (N), P, and Potassium (K). The type of manure spread each year on a given field was randomly set to a given type with an overall proportion of 30%, 55% and 15%, respectively for cattle, pig and poultry slurries. Two fertilization scenarios were successively implemented:

- N-based fertilization during the first 20 years: the amount of organic manure spread on a given field is derived from crop needs in N, with an upper limit of 170 kg N/ha fixed by regulation. P loads were then estimated using the P/N ratios of the different slurries (Table A.I.12.a);
- P-based fertilization during the following 30 years: the amounts of organic manure spread on a given field is derived from crop needs in P, soil P content and fertilization background (Schvartz et al., 2005).  $P_{\text{reinforced}}$  is the soil P content below which it is recommended to increase the P load spread on the field.  $P_{\text{skip}}$  is the soil P content above which it is reasonable not to spread manure on the field, although the method forbids skipping fertilization two years in a row. We used the numerical values recommended on the actual region (Brittany) of the studied site (Table A.I.12).

Both fertilization scenarios were applied at field-scale to reflect the management units of the farmers, where the soil P contents used were calculated from the mean value of the pixels within each field. The switch of fertilization practices accounted for the implementation of a new legislation concerning best management practices and by a change of perception in the agricultural community towards environmental issues.

## Soil P budget

Inputs of the soil P budget model were P loads from fertilizers, outputs are crop P offtakes and a fraction of the soil extractable P which annually turns into a non-extractable form. The soil P stock of a given pixel  $i$  at the end of year  $y$  can be calculated as follows:

$$P_{i,y} = (1-k)P_{i,y-1} + Pf_{i,y} - PE_{i,y} \quad (8)$$

where  $P_{i,y}$  is the extractable soil P content at year  $y$  of pixel  $i$ ,  $k$  is the proportion of extractable soil P which turns into a non-extractable form,  $Pf_{i,y}$  is the P load input by P fertilization during year  $y$  on pixel  $i$ , and  $PE_{i,y}$  is the extractable P exported from pixel  $i$  by the crop harvested on year  $y$ . Phosphorous exports were calculated at the pixel scale following the estimation of crop P needs modified with a stochastic factor:

Sample size	Sampling density	Sampling designs			
		<i>square regular grid</i>	<i>stratified on initial landuse</i>	<i>stratified on current landuse</i>	<i>8 points transect</i>
(sites)	(site/km <sup>2</sup> )	(size of the mesh, m)	(points by landuse)	(points by landuse)	(number of transects)
16	0.1	3000	4	4	2
144	0.6	1150	36	36	18
256	1.1	880	64	64	32
576	2.6	600	144	144	72
1024	4.6	450	256	256	128

Table A.I.13: Sample sizes and their implementation within the different strategies.

$$PE_{i,y} = P_{export} Y_{crop} = P_{export} N(\text{mean}_{\text{yieldcrop}}, \text{std}_{\text{yieldcrop}}) \quad (9)$$

where  $P_{export}$  is the crop P ratio, and  $Y_{crop}$  is the crop yield (Schvartz et al., 2005) derived from a normal distribution characterized by a mean yield value and a yield standard deviation set to 10% of the mean (Table A.I.12.b).

## **Sampling strategies**

We kept one realization of the soil P spatio-temporal model to implement different sampling strategies, which combined four sampling designs, five sample sizes, and two time intervals. We used the sample median to estimate the landscape median (the median value derived from the whole landscape), a robust estimator in situation where the distribution of the surveyed variables may be unknown. Each sampling strategy was repeated 100 times for statistical purposes.

### **Sampling designs**

We tested the following sampling designs:

- regular grid (RG) is an aligned systematic sampling in two dimensions (Cochran, 1977). For a square grid of  $n$  points by side, the sidelength of the landscape was divided in  $n$  segments and the starting site of the grid was randomly allocated within the square defined by the first longitudinal and latitudinal segments. The other sites were allocated by translating the location of the first site with the sidelength mentioned above (see mesh sizes on Table A.I.13). Allocated the first year, sample sites stayed stationary during the remaining 50 years;
- stratified by initial landuse (IL): the landscape area can be divided into four strata based on the four different landuses considered here. At the first year, a simple random sampling without replacement was performed on each landuse stratum, each stratum being given the same number of sampling sites. Allocated the first year, sample sites stayed stationary during the remaining 50 years;
- stratified by current landuse (CL): same process as above, but all the sites were reallocated at each timestep based on the landuse of the year of interest;
- transects (Tr): a transect is a line of sampling sites separated by a short distance. In this design, each transect was a line of eight consecutive pixels. Starting points of transects were randomly allocated and then, the direction of each transect was randomly set among the four possibilities: East-West, North-South, North East-South West, and North West-South East. The distance between the consecutive sites of a transect was, in our study, the distance between consecutive pixels: 50 m for EW and NS directions, 71 m for NE-SW and NW-SE directions. Allocated the first year, sample sites stayed stationary during the remaining 50 years.

RG and Tr are self-weighting designs, therefore the sample median is an unbiased estimator of the landscape median (for an even sample size, the sample median was the mean of the lower and upper medians). For IL and CL designs, we used the median unbiased estimator for stratified samplings suggested by Lohr (1999).

Each sample unit  $i$  is associated to a weight  $w_i$ , which is the inverse of its probability of being sampled. The rasterization of the virtual landscape in 50m by 50m pixels implies that only a finite number of sites can be sampled. Thus, in a stratified sampling, the probability to sample a given pixel  $i$  located in strata  $h$  is the number of sampling sites allocated to strata  $h$  divided by the total number of pixels of strata  $h$ . The *empirical distribution function*  $\hat{F}(y)$  is the sum of all weights for sample values  $y_i$  so that  $y_i \leq y$ , divided by the sum of all the weights carried by the sampling sites  $i$  of sample  $S$ :

$$\hat{F}(y) = \frac{\sum_{i \in S: y_i \leq y} w_i}{\sum_{i \in S} w_i} \quad (10)$$

The median can then be estimated by taking the smallest sample value  $y_i$  for which  $\hat{F}(y_i) \geq 0.5$  (Gross, 1980). The weights of the sample sites for IL samples referred to the total number of pixels of each landuse stratum at the first year and then remained constant, whereas the weights for the CL samples were computed each year with the total number of pixels of the current landuse strata.

### **Sample sizes and sampling frequency**

Five sample sizes were implemented (Table A.I.13) to observe the effect of sampling density on sampling design accuracy and trend detectability. We considered a two- and an eight-year interval between sampling campaigns to study the effect of the sampling frequency on sampling design accuracy and ability to detect trends.

### **Evaluation of sampling strategies**

Each year, the sample median of each realization of each strategy was stored as an estimate of the landscape median value. We used different criteria based on the comparison of the landscape and the sample medians to assess the accuracy and the ability to detect evolution trends of the strategies described beforehand.

### **Sampling design accuracy**

We used two statistics to assess the accuracy of each design:

- the bias of the median estimate:

$$bias = \frac{1}{N_{Year}} \frac{1}{N_{Realization}} \sum_{y=1}^{N_{Year}} \left[ \sum_{r=1}^{N_{Realization}} (Med_{y,r} - Med_{y,L}) \right] \quad (11)$$

where  $N_{Year}$  is the number of years when a sampling occurred,  $N_{Realisation}$  is the number of repeated realizations of each strategy,  $Med_{y,r}$  is the sample median for the sample  $r$  at year  $y$ , and  $Med_{y,L}$  is the landscape median at year  $y$ .

- the root mean square error (RMSE) of the median estimate:

$$RMSE = \sqrt{\frac{1}{N_{Year}} \frac{1}{N_{Realization}} \sum_{y=1}^{N_{Year}} \left[ \sum_{r=1}^{N_{Realization}} (Med_{y,r} - Med_{y,L})^2 \right]} \quad (12)$$

Biases and RMSE integrated all the realizations of each sampling strategy and were calculated over the whole duration of the simulation.

### Trend detection

We designed a trend ratio (TR) criterion to assess both the trend detection ability and its accuracy for a given design:

$$TR = \frac{(Med_{y,r} - Med_{y-t,r})}{(Med_{y,L} - Med_{y-t,L})} \quad (13)$$

where  $t$  is the time step in years between two sampling dates. To avoid a division by zero, a logical test was performed:

$$\begin{aligned} &\text{if } Med_{y,A} \neq Med_{y-t,A} \text{ then} \\ &\quad TR = \frac{(Med_{y,r} - Med_{y-t,r})}{(Med_{y,L} - Med_{y-t,L})} \\ &\quad \text{else if } Med_{y,r} = Med_{y-t,r} \text{ then} \\ &\quad \quad TR = 1 \\ &\quad \text{otherwise} \\ &\quad \quad TR = 0 \\ &\text{endif} \end{aligned} \quad (14)$$

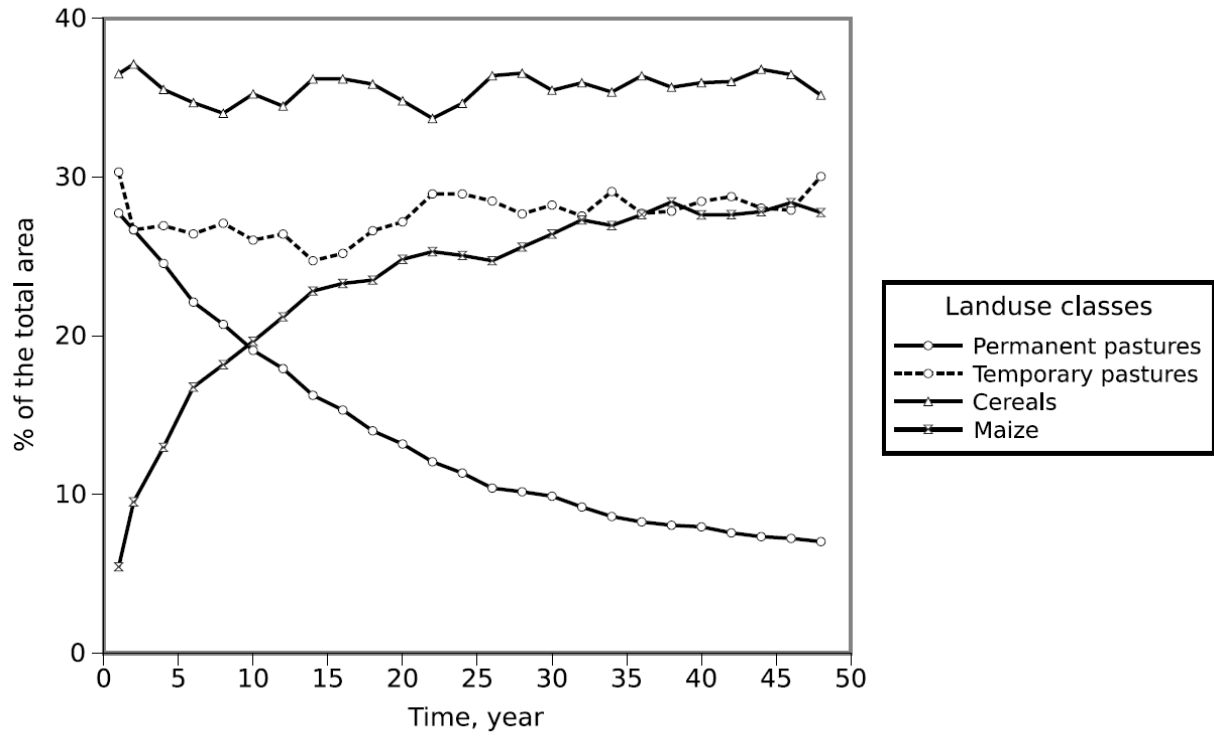


Fig. A.I.36: Simulated evolution over 50 years of the proportional area occupied by each landuse class.

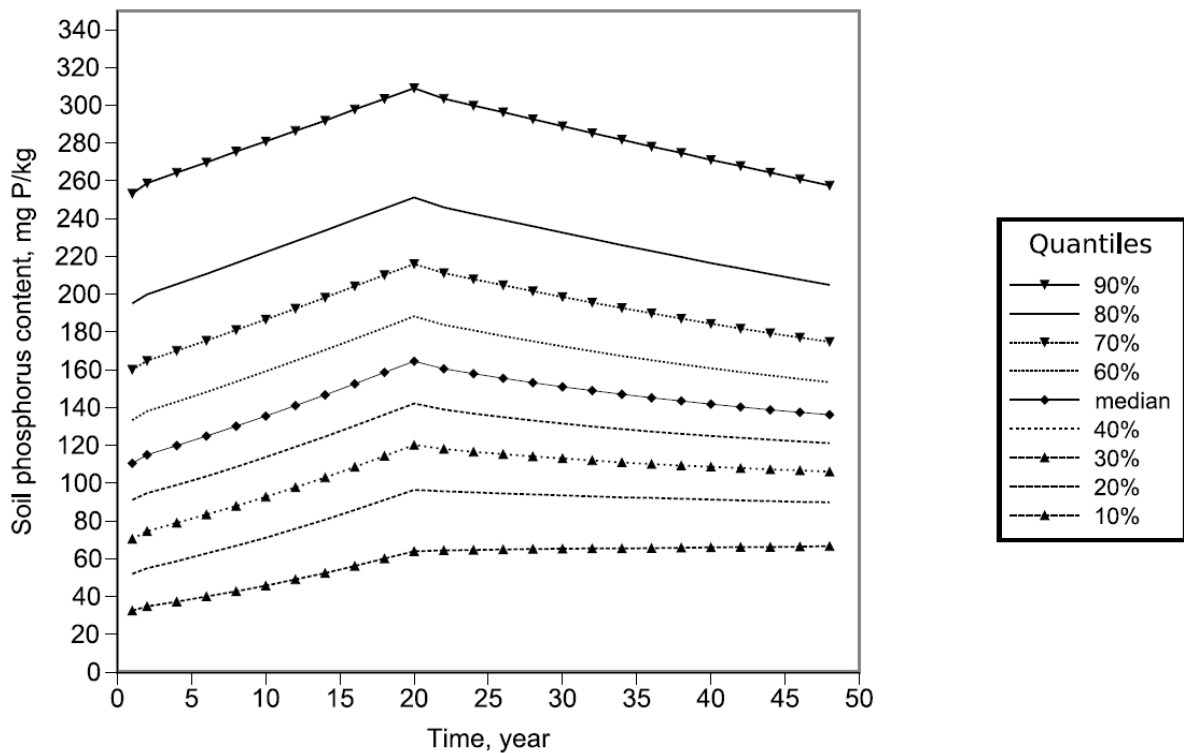


Fig. A.I.37: Evolution over 50 years of the soil P distribution over the whole landscape.

The rules we followed to interpret this criterion were:

- $TR > 0$  when the soil P temporal trend detected by the design (called the sample evolution) follows the trend observed on the virtual landscape (called the landscape evolution);
- $TR < 0$  when the sample evolution is the opposite to the landscape evolution;
- $TR = 0$  when the sample median value stays constant. If both the landscape and the sample median value stay constant,  $TR = 1$ . If only the landscape median value is constant,  $TR$  is set to 0 because there is no trend which should be detected by the sample median. These cases will seldom occur;
- as a ratio, the closer  $TR$  is to one, the better is the estimation of the evolution by the design.

Therefore, the  $TR$  ratio allows the assessment of trend detection ability and accuracy. A  $TR$  value would indicate a successful trend detection if and only if landscape and sample temporal trends evolve in the same way, which can be formalized with the following condition:

$$TR \geq 0 \quad (15)$$

One can expect a limited level of bias between the landscape temporal trend and the sampled one. Defined as a ratio between the sample and the landscape trends,  $TR$  can be framed between two thresholds to avoid both under- and over-estimation by the sample trend. In our numerical application, we arbitrarily defined accurate trend detection with the following conditions:

$$\frac{1}{2} \leq TR \leq 2 \quad (16)$$

## Results

The evolution of the landuse proportions is shown on Fig. A.I.36. The transition probabilities reproduced the change of landuse which occurred in Brittany from the 1960's until now, namely a diminution of permanent pastures from 30 to 7% of the total area, globally replaced by maize fields, and a rather constant area of other cereals and temporary pastures (36% and 27%, respectively).

### ***Initial state of the soil P content distribution***

Fig. A.I.37 shows the evolution of soil P content distribution. At year 1, 25%-median-75% quartile values of the virtual landscape (61, 110.5, and 178 mg P/kg) are similar to the BDAT quartiles (63, 112, and 179 mg P/kg), due to the backward NST.



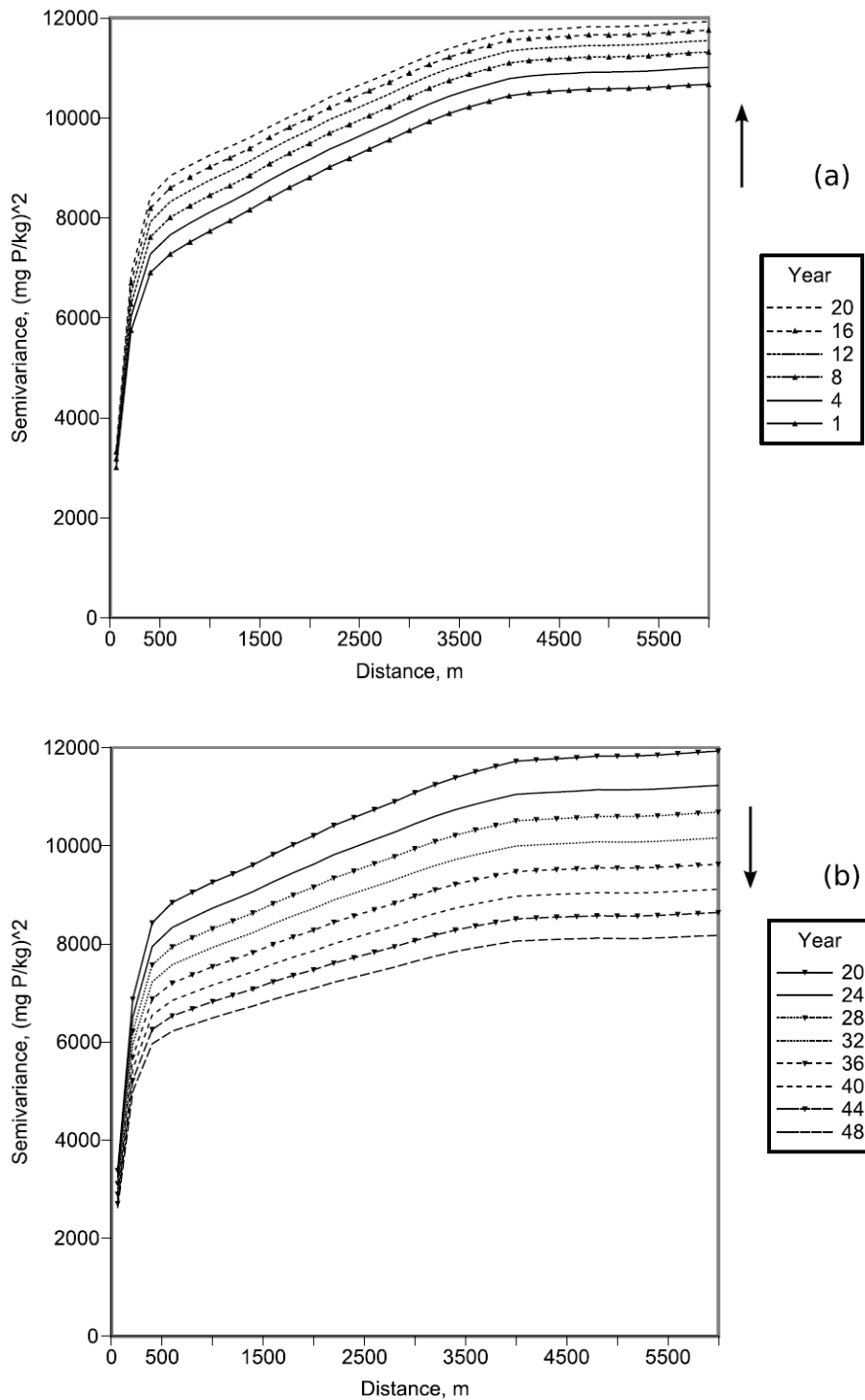


Fig. A.I.38: Evolution of the soil phosphorus (P) content spatial structure during the two phases of the simulation: (a) soil P variogram evolution between year 1 and 20 (N-based fertilization practices); (b) soil P variogram evolution between year 20 and 48 (P-based fertilization practices).

The sill of the experimental variogram at year 1 (Fig. 5) was similar to the sill of the BDAT variogram (9500 (mg P/kg)<sup>2</sup>). The underlying short-, medium-, and long-range spatial structures are observable through abrupt changes of slope at 300 m, 600 m, and 4000 m.

## ***Global evolution of soil P content***

### **Evolution of the distribution**

During the N-based fertilization period, all the decile curves, except the 10% decile, stayed parallel to the landscape median evolution curve (Fig. A.I.37). The shape of the distribution was not modified for medium and high values, but translated towards higher values due to a P over-fertilization. The 10% decile mainly accounts for soil P content in pastures, which had a low initial soil P content (medium-range component), were less fertilized than the other crops, and had a high probability to remain as pastures (Table 11).

After 20 years, sustainable P-based fertilization practices gradually smoothed the soil P content spatial distribution shape through time, by raising low P content values and reducing high P content values. While the median value decreased, higher deciles values decreased with a higher rate, and lower deciles decreased with a lower rate, thus narrowing the scope of the distribution and translating its mode value above the  $P_{\text{reinforced}}$  value (100 mg P/kg). This behaviour illustrated the tendency of this P-based fertilization to homogenize soil P content to a value which is considered agronomically optimal.

### **Evolution of the spatial structure**

Besides the distribution modifications, fertilization practices also had an impact on the spatial structure of the soil P content landscape. Fig. A.I.38.a and b show the variogram evolution over respectively the first and the second simulation period. In both cases, the short-range component hardly evolved because the manure load was applied uniformly at field scale, thus not modifying the intra-field variability. However, medium- and long-range variability successively increased and decreased. This is mainly due to the definition of the fertilization practices. Fig. A.I.38.a shows an increase of the medium- and long-range components. The N-based fertilization applied smaller P loads on pastures and heavier P loads on other crops. Moreover, because of the low probability for a cereal or a maize crop to turn back to a permanent pasture, the soil P content of permanent pastures tended to increase more slowly than cereal or maize fields, thus increasing the variance at the medium- and large-range scales. However, the P-based fertilization practices tended to smooth soil P content values (Fig. A.I.38.b), at the field scale, towards an agronomically optimal value, thus decreasing the variance at the medium- and large-range scale.

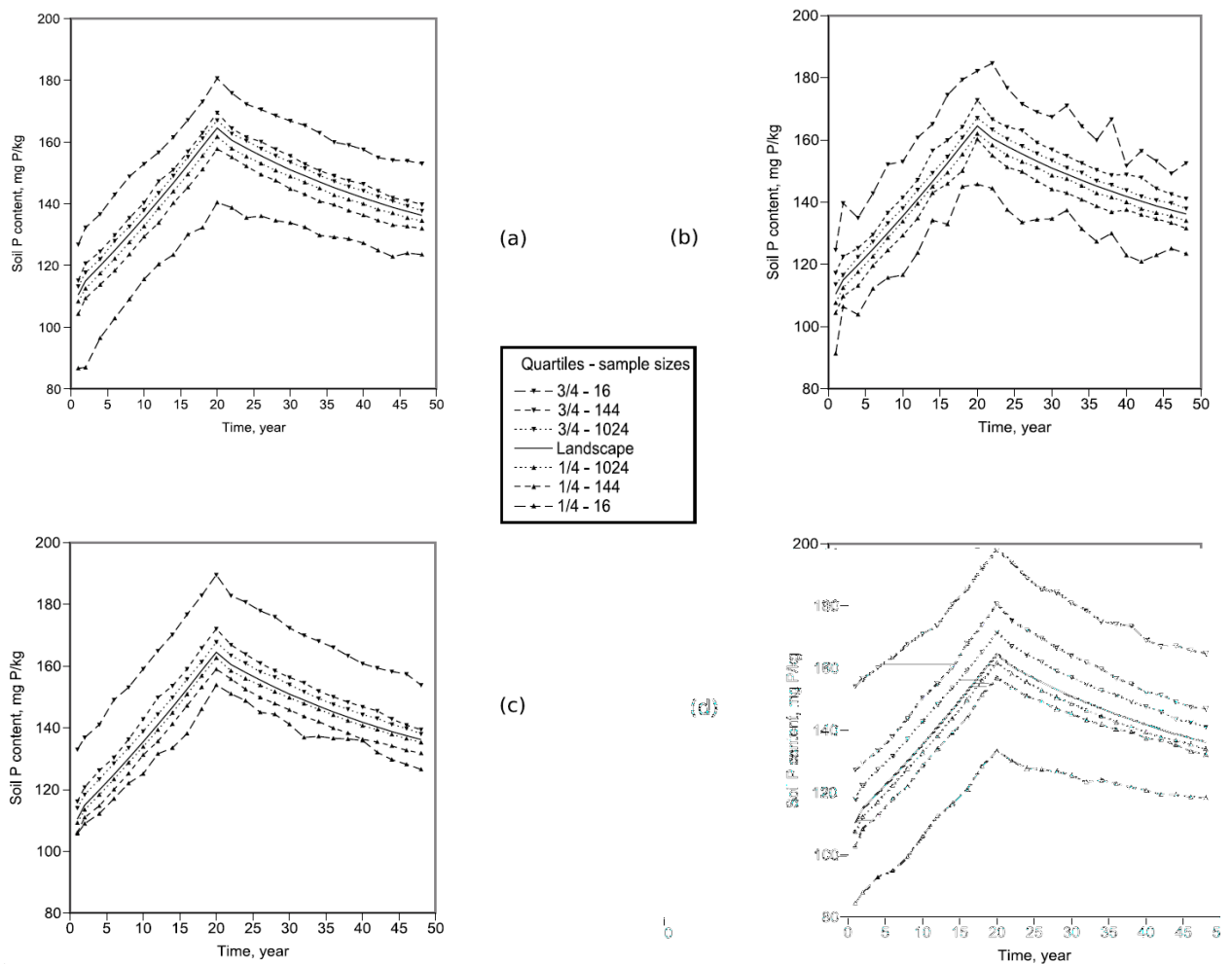


Fig. A.I.39: Evolution of the landscape median of soil P content surrounded by the dispersion envelopes (DE) of the sampling strategies. The DE, based on the first and third quartiles of 100 realizations of each sample, are represented for three sample sizes (16, 144, and 1024) and the four sampling designs implemented in this study: (a) initial landuse; (b) current landuse; (c) regular grid; (d) transect.

## Strategy evaluation

### Accuracy assessment

Fig. A.I.39 shows the temporal evolution of the landscape median surrounded by the dispersion envelopes (DE) derived from the 1<sup>st</sup> and 3<sup>rd</sup> quartiles of the 100 repeated realizations of each sample, for the following sample sizes: 16, 144, and 1024. The range of the decreased with increasing sample size, particularly between 16 and 144 sampling sites. Transect design was characterized by the widest DE. The DE curves of stationary designs (IL, RG, and Tr) appeared centered on the landscape median curve, whereas the non-stationarity of CL sample sites reflected in scattered DE curves, particularly for low sample sizes.

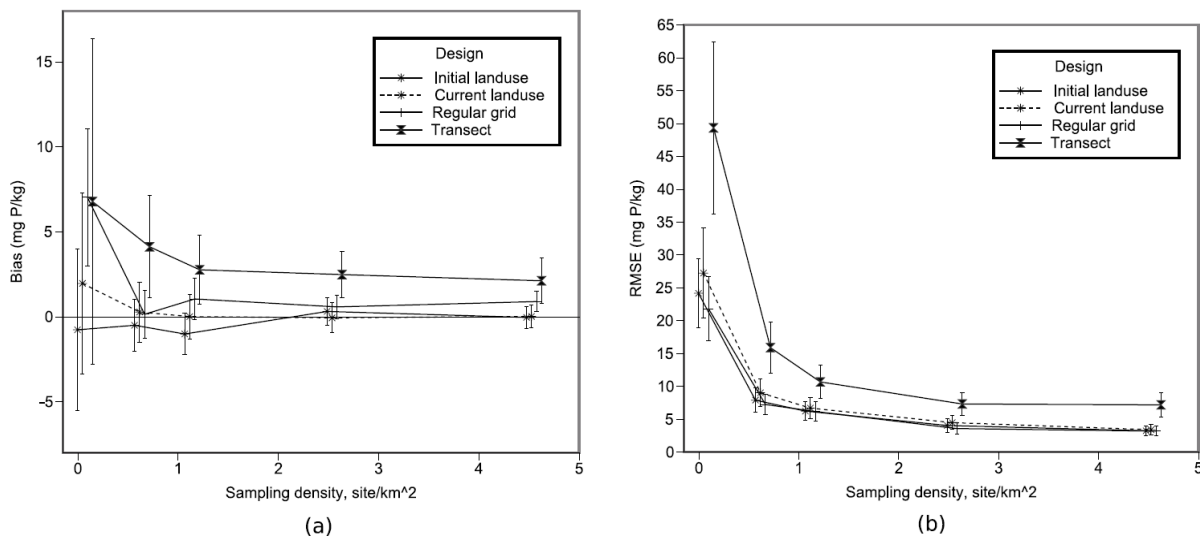


Fig. A.I.40: Accuracy of the sampling designs as a function of sampling density: (a) mean bias; (b) mean RMSE; along with their 95% confidence intervals. X-coordinates were slightly jittered to facilitate the reading of the confidence intervals.

Fig. A.I.40 represents the accuracy of the sampling strategies by design and sampling density:

- between 0.1 and 1.1 site/km<sup>2</sup>, biases and RMSE of all the strategies dropped and confidence intervals (CI) narrowed with an increase of the sampling density. Sampling densities above 1.1 site/km<sup>2</sup> did not significantly improve neither biases, RMSE nor CI;
- both Tr and RG samples exhibited some positive bias, despite their self-weighting designs. Transect-based samples showed the widest CI for bias and RMSE, whereas RG samples showed low RMSE and narrow CI, even for low sample densities;

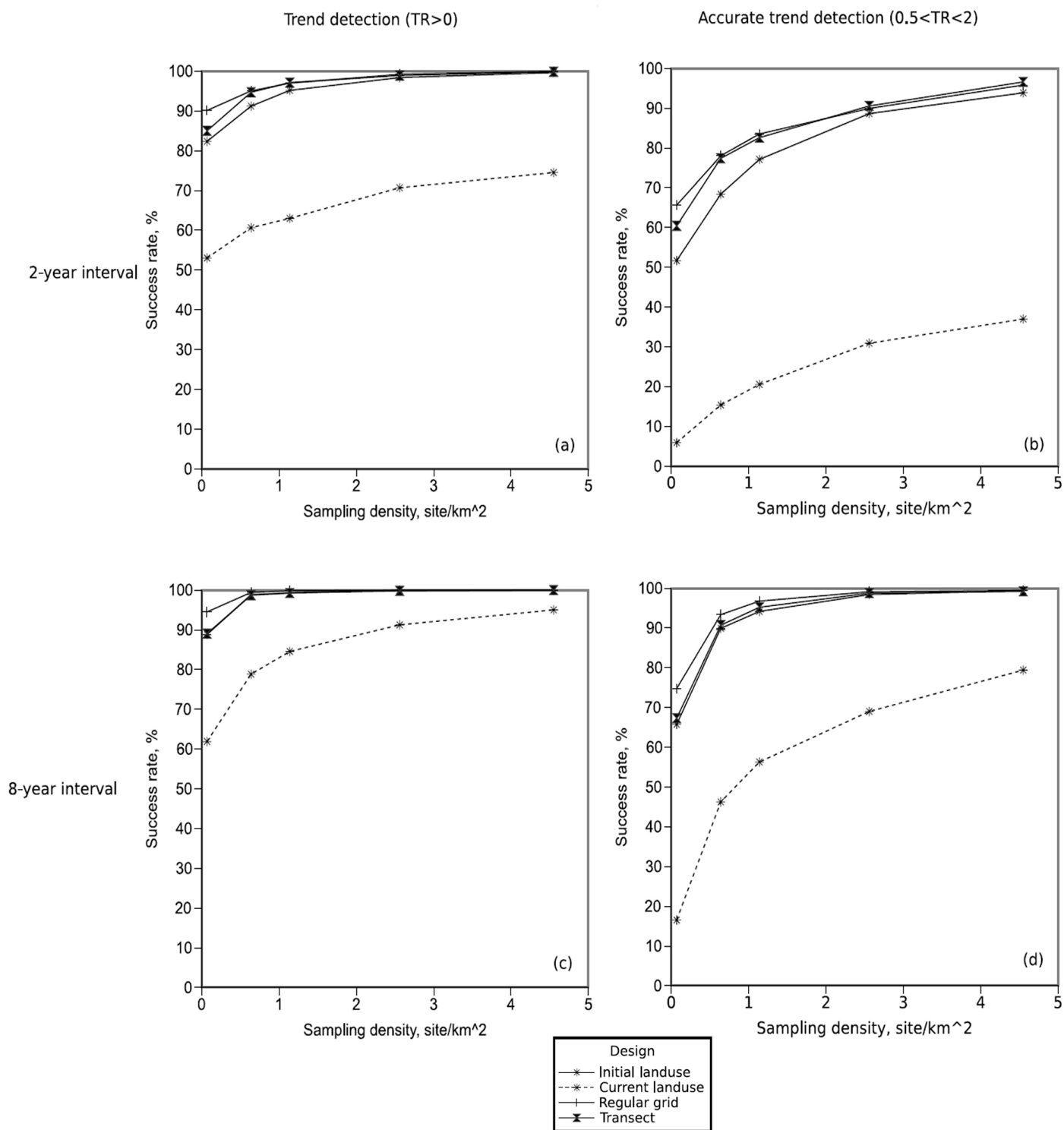


Fig. A.I.41: Evolution of the detection rate as a function of sampling density for the different sampling designs and sampling time-intervals. Trend detection for a 2 years (a) and 8 years (c) sampling time-interval; accurate trend detection for a 2 years (b) and 8 years (d) sampling interval.

- samples stratified by landuse (IL and CL) exhibited less bias than RG and Tr. IL was the only design to be negatively biased (although not significantly, considering the CI). During the landscape evolution process, the size of the permanent pastures stratum decreased from 28% to 7% of the total area. Therefore, the weights initially associated with the sampling sites located in permanent pastures slightly over-represented the low soil P content values carried by permanent pastures as the landscape evolved (see the cultural crop components in Table A.I.10).

Finally, samples based on IL and CL designs were the most accurate in terms of bias. RG, IL, and CL designs were significantly better than Tr in terms of RMSE. Tr was the worst strategy for all sample sizes, in terms of DE, bias, and RMSE.

### **Trend detectability**

The trend ratio (TR) was developed to assess trend detectability of sampling strategy. The approaches used here were based on a success rate. Each pair of sample median values separated by a time interval within a realization was characterised by a TR value. The success rate is the ratio of satisfying TR values to the total number of pairs, which depended on the considered time interval. Fig. A.I.41 shows the effect of sampling density and frequency on the trend detectability of the sampling strategies. A higher sampling density increased the success rates of all the strategies for both 2- and 8-year time intervals, and particularly for sampling densities comprised between 0.1 and 1.1 site/km<sup>2</sup> where success rates gained from 5 to 40%. The success rate improvements were non-linear with sampling densities and tended asymptotically toward 100%.

Considering only trend detectability (TR>0), stationary designs had high success rates (greater than 85%), either with a 2-year or a 8-year interval. Despite its good accuracy, the trend detectability of the CL design was clearly disadvantaged by the relocation of the sampling sites: at a 2-year interval, the success rate was still below 75% with a sampling density of 4.6 site/km<sup>2</sup>. However, success rates of CL could be significantly improved by increasing the time interval between samplings to 8 years because the soil P content landscape was in a context of long term and uniform evolution (20 years of intensive fertilization and then 30 years of balanced fertilization): the success rate rose from 60% to nearly 80% for a sampling density of 0.6 site/km<sup>2</sup>.

The framing of the trend ratio ( $0.5 < TR < 2$ ) decreased the success rates of all strategies, mainly for low sampling densities: -40% for CL at 0.1 site/km<sup>2</sup>, -25% for the others at the same sampling density. The accurate approach showed a stronger influence of sampling density on trend detection: for a two-year interval, the success level gain between 0.1 and 4.6 sites/km<sup>2</sup> was around 15% for absolute detection trend, and 35% for accurate trend detection; for an eight-year interval and the CL design, the same success level gain was around 30% for absolute detection trend, and 70% for accurate trend detection.

The trend detection assessment confirmed the best performance of stationary designs (RG, Tr, and IL), which had success rates significantly higher than the dynamic CL design.

## Discussion and conclusion

An agricultural landscape was virtually constructed by combining stochastic and deterministic components, which had complementary spatial scales. Its evolution was modelled by a soil P budget driven by landuse, fertilization practices, crop exportations and soil immobilization. During the N-based scenario, the soil P content median of the landscape increased at a mean rate of  $2.8 \text{ mg P kg}^{-1} \text{ year}^{-1}$ , which is consistent with observed rates of the  $2.7 \text{ mg P kg}^{-1} \text{ year}^{-1}$  rate observed in Brittany between 1980 and 1999 within the BDAT (Lemerrier et al., 2006). The same study showed that the interquartile difference remained constant over the same period of time ( $120 \text{ mg P kg}^{-1}$  in the virtual landscape;  $118 \text{ mg P kg}^{-1}$  observed in the BDAT). To our knowledge, survey data on agricultural areas with P-based fertilization practices over several decades are not available yet to validate the second phase of the landscape evolution. Although the degree of realism of such a landscape is hard to evaluate because a comparison with sufficient spatio-temporal data is hardly affordable (Walter et al., 2003), this simulation approach provides a flexible way to design a large range of virtual landscapes, either by modifying some characteristics (field pattern, landuses, and spatial distributions) or the evolutionary processes (landuse changes, fertilization practices, and budget calculations).

In this study, the virtual landscape enabled to assess the ability of survey plans to detect the effects on soil P content of changes in agricultural management practices. We demonstrated that the lack of local data can be compensated by the implementation of known spatial structures and known evolution processes. We addressed three classical parameters of sampling strategies to evaluate our methodology based on virtual landscape: the sampling design, the sampling density and the time interval between two sampling campaigns. The results concerning the influence of sampling densities and the ability of the designs to detect evolution trends follow the conclusions of other studies (Papritz et al., 1995). Nevertheless, we demonstrated that the self-weighting designs implemented here (RG and Tr) on a spatially structured and evolving agroecosystem exhibited significantly more bias than designs stratified by landuse. In this study, permanent pastures were characterized by low P content values and were located preferentially in the steep areas that a regular sampling design may not sample representatively. On the other hand, we showed that IL design was not biased, despite its stratification on the initial state of a dynamic variable (landuse), on which fertilization practices were based. This may be related to the cultural crop component of the initial soil P content distribution, which accounted for the effect of fertilization history of landuses like permanent pastures and maize. Moreover, Fig. A.I.39 shows that the accuracy of IL does not decrease with time, at least not within the duration of the simulation. Therefore, a design stratified by the initial state of a dynamic variable can be a recommended choice for survey designers, when it is known that the chosen state accounts for a specific history, thus inducing long-lasting sample weights. In our context, IL stood for the most appropriate sampling design, although there was no absolute ranking of the tested designs (see Table 14).

	<i>best design</i>			<i>worst design</i>
BIAS	current landuse	initial landuse	regular grid	transect*
ERROR	regular grid	initial landuse	current landuse	transect*
TR success rate	regular grid	transect	initial landuse	current landuse*

Table A.I.14: Sampling design ranking in terms of accuracy and trend detection. Designs marked with a \* showed significant difference of performance.

We showed that a shift in fertilization practices implied changes of the P distribution and spatial structure. To account for the anisotropic or non-stationary spatial structures, previous work suggested to modify the sampling design during the survey (Lesser et al., 1997) or to implement a dynamic model-based sampling design when the spatio-temporal model of the surveyed variable is known (Wikle et al., 2005). Model-based approaches optimize the design and the sample size of a strategy by modeling the spatial structure of the studied variable (McBratney et al., 1983). Our work also raises the perspective of combining model-based and sampling-based approaches to implement a multi-stage sampling approach, in which the variogram deduced from the preceding sampling campaign would be used to move some of the sampling sites in order to improve the subsequent sampling.

The modeling of the effect of agricultural practices on the environment can be improved in two manners. Firstly, P leaching and runoff could be introduced into the soil P budget model because such fluxes have strong impacts on water quality (Sharpley et al., 2001). Secondly, instead of being managed independently, the fields should be aggregated into their original farm system to account for known landuse spatial patterns and technical constraints characterizing the main production of the farm (Thenail et al., 2004) and fertilization practices. Instead of the probabilistic rules used to define both landuse changes and the origin of manures, deterministic rules associated to the farm types represented in the landscape could integrate other driving constraints such as: distance between pastures to the farmstead, manure composition depending on cattle production, crop affinity to soil hydromorphy, etc. The transition probability matrix was constant over time and though it accounted well for a global landuse change, its lack of determinism may cause local discrepancies in the landuse of a farm system.

Although we focused on soil P content, other environmental variables (like organic carbon, N or K content, various soil hydraulic properties, land use characteristics) are very likely to be monitored for their agronomic interest at the same time, thus adding a new dimension to an assessment study. This raises the challenge to be able to sample efficiently every monitored parameter and to find an optimal compromise of the sample size. Moreover, a significant part of the cost of a sampling program is due to the preparation or the maintenance of sample sites. Therefore, a maximum number of variables of interest may be monitored at each site to make the survey cost-efficient. The more multidimensional a survey program becomes, the more attention survey designers must pay to avoid every sources of bias for one or many of the measured variable. Simulations on sufficiently realistic virtual landscapes should help to test different sampling strategies for multivariable monitoring on their accuracy and their ability to detect trends of evolution.



## References

- Anderson A.B. 2002. Detecting changes in natural resources using land condition trend analysis data. *Environmental Management*, 29:428-436.
- Aurousseau P. 2001. Les flux d'azote et de phosphore provenant des bassins versants de la rade de Brest. Comparaison avec la Bretagne. *Oceanis*, 27 (2):137-161.
- Boulaine J. 1996. Histoire de l'Agronomie en France. Lavoisier Tec&Doc.
- Brus D.J., and J.J. de Gruijter. 1997. Random sampling or geostatistical modelling? Choosing between design-based and model-based sampling strategies for soil (with Discussion). *Geoderma*, 80:1-44.
- Carriquiry, A.L., F.J. Breidt, P.G. Lakshminarayan. 1998. Sampling schemes for policy analyses using computer simulation experiments. *Environmental Management*, 22 (4):505-515.
- Cochran W.G. 1977. Sampling techniques (3rd edition.). John Wiley & Sons, New York
- Conant R.T., G.R. Smith, K. Paustian. 2003. Spatial variability of soil carbon in forested and cultivated sites: implications for change. *Journal of Environmental Quality*, 32 (1):278-286.
- Cressie N.A.C. 1991. Statistics for spatial data. John Wiley & Sons, New York.
- De Gruijter J.J., D.J. Brus, M.F.P. Bierkens, M. Knotters. 2006. Sampling for natural resource Monitoring. Springer-Verlag, Berlin.
- Department for Environment, Food and Rural Affairs. 2004. The first soil action plan for England:2004- 2006.
- Dougherty, W.J., N.K. Fleming, J.W. Cox, and D.J. Chittleborough. 2004. Phosphorus transfer in surface runoff from intensive pasture systems at various scales: a review. *Journal of Environmental Quality*, 33:1973-1988.
- Fardeau J.C. and A.P. Conesa. 1994. Le phosphore. In Bonneau M. & Souchier, B. *Pédologie 2. Constituants et propriétés du sol*:557-568.
- Fernández J.A., C. Real, J.A. Couto, J.R. Aboal, A. Carballeira. 2005. The effect of sampling design on extensive bryomonitoring surveys of air pollution. *Science of the Total Environment*, 337:11-21.
- Goovaerts P. 1997. Geostatistics for natural resources evaluation. Oxford Univ. Press, New York.
- Gross S.T. 1980. Median estimation in sample surveys. In Proceedings of the Survey Research Methods Section. Alexandria, Virginia. American Statistical Association:181-184.
- Heathwaite A.L., A.I. Fraser, P.J. Johnes, M. Hutchins, E. Lord, D. Butterfield. 2003. The Phosphorus Indicators Tool: a simple model of diffuse P loss from agricultural land to water, *Soil Use and Management*, 19:1-11.
- Kaluzny S.P., S.V. Vega, T.P. Cardoso, A.A. Shelly. 1998. S+ Spatial Stats User's Manual, Springer (ed.).
- King D., P. Stengel, M. Jamagne, C. Le Bas, D. Arrouays. 2005. Soil mapping and soil monitoring: state of progress and use in France. In: *Soil Resources of Europe*, second edition. R. Jones, B. Houskava, P. Bullock, and L. Montanarella (eds). European Soil Bureau Research Report n° 9. EUR 20559 EN. Office for Official Publications of the European Communities, Luxembourg:139-146.
- Lemercier B., C. Walter, C. Schwartz, N. Saby, D. Arrouays, S. Follain. 2006. Organic carbon and extractable phosphorus contents monitoring in cultivated soils of three French regions. *Étude et Gestion des Sols*, 73 (3):165-179.
- Lesser V.M. and W.D. Kalsbeek. 1997. A comparison of periodic survey designs employing multi-stage sampling. *Environmental and Ecological Statistics*, 4:117-130.
- Lohr, S.L. 1999. Sampling: design and analysis. Brooks/Cole Publishing, Pacific Grove, California.
- McBratney A.B. and R. Webster. 1983. How many observations are needed for regional estimation of soil regional properties? *Soil Science*, 135:177-183.
- McDowell R.W, and A.N. Sharpley. 2001. Approximating phosphorus release from soils to surface runoff and subsurface drainage. *Journal of Environmental Quality*, 30:508-520.
- Minasny B., D.J.J. Walvoort, A.B. McBratney. 2007. The variance quadtree algorithm: use for spatial sampling design. *Computers & Geosciences*, 33 (3):383392.
- Némery J., J. Garnier, C. Morel. 2005. Phosphorus budget in the Marne Watershed (France): urban vs. diffuse sources, dissolved vs. particulate forms. *Biogeochemistry*, 72:35-66.

- Papritz A. and R. Webster. 1995. Estimating temporal change in soil monitoring: II. Sampling from simulated fields, *European Journal of Soil Science*, 46:13-27.
- Saby N. and D. Arrouays. 2004. Simulation of the use of a soil-monitoring network to verify carbon sequestration in soils: will changes in organic carbon stocks be detectable? *Soil Science and Plant Analysis*, 35 (17-18):2379-2396.
- Schvartz C., J.C. Muller, J. Decroux. 2005. Guide de la fertilisation raisonnée. France Agricole (ed.).
- Sharpley A.N., R.W. McDowell, P.J.A. Kleinman. 2001. Phosphorus loss from land to water: Integrating agricultural and environmental management. *Plant Soil*, 237:287–307.
- Simbahan, G.C., and A. Dobermann. 2006. Sampling optimization based on secondary information and its utilization in soil carbon mapping. *Geoderma*, 133 (3-4): 345-362.
- Sprent P. 1989. *Pratique des statistiques non-paramétriques*. Traduction Ley, J.-P., INRA (ed.).
- Thenail, C. and J. Baudry. 2004. Variation of farm spatial land use pattern according to the structure of the hedgerow network (bocage) landscape: a case study in northeast Brittany. *Agricultural Ecosystems & Environment*, 101:53-72.
- Vighi, M., and G. Chiaudani. 1987. Eutrophication in Europe: the role of agricultural activities. 213-257 In Hodgson, E. (ed.) *Reviews in Environmental Toxicology* 3, Elsevier.
- Walter C., C. Schvartz, B. Claudot, T. Bouédo, P. Arousseau. 1997. Statistical review of the soil tests made in France from 1990 to 1994: statistical and cartographic descriptions of the cultivated topsoil horizon variability. *Étude et Gestion des Sols*, 4 (3):191-204.
- Walter C., R.A. Visccarra Rossel, A.B. McBratney. 2003. Spatio-temporal simulation of the field-scale evolution of organic carbon over the landscape. *Soil Science Society of America*, 67:1477-1486.
- Wikle C.K. and J.A. Royle. 2005. Dynamic design of ecological monitoring networks for non-Gaussian spatio-temporal data. *Environmetrics* 16, (5):507-522.



## **Annexes II**

Matrices de transition et arbres de décision stochastiques  
relatifs aux résultats des chapitres III et IV



# Annexes II : Matrices de probabilités de transition et arbres de décision stochastiques complétant les résultats du chapitre III et IV

## Matrices de transition de cultures pour les différents types d'exploitation du bassin versant de Naizin

Les matrices de transition suivantes ont été établies par *Rotomatrix* (Chapitre III) pour l'ensemble du bassin versant (*generic*) et par type d'exploitation (*specific*), à partir d'un jeu d'apprentissage constitué de l'inventaire des cultures sur la période 1993-1998. Les matrices indiquent la probabilité de passage d'une culture en ligne présente sur une parcelle l'année  $n$  à une culture en colonne présente l'année  $n+1$ .

generic landuse N	landuse N+1									
	<i>fallow</i>	<i>grain</i>	<i>maizegrain</i>	<i>maizesilage</i>	<i>other</i>	<i>permpasture</i>	<i>potato</i>	<i>raygrassseed</i>	<i>temppasture</i>	<i>vegetable</i>
<i>fallow</i>	0.333	0.268	0.133	0.133						0.133
<i>grain</i>	0.006	0.2	0.331	0.146	0.006		0.041	0.006	0.128	0.136
<i>maizegrain</i>	0.015	0.466	0.324	0.018			0.024		0.033	0.12
<i>maizesilage</i>	0.015	0.404	0.033	0.241	0.006		0.006		0.202	0.093
<i>other</i>		0.4	0.3				0.1			0.2
<i>permpasture</i>						1				
<i>potato</i>		0.377	0.361	0.098					0.082	0.082
<i>raygrassseed</i>		0.167	0.25					0.583		
<i>temppasture</i>	0.005	0.079	0.032	0.117	0.003		0.005		0.709	0.05
<i>vegetable</i>	0.008	0.452	0.14	0.136		0.021	0.033		0.103	0.107

dairy landuse N	landuse N+1									
	<i>fallow</i>	<i>grain</i>	<i>maizegrain</i>	<i>maizesilage</i>	<i>other</i>	<i>permpasture</i>	<i>potato</i>	<i>temppasture</i>	<i>vegetable</i>	
<i>grain</i>		0.234	0.078	0.319			0.015	0.326	0.028	
<i>maizegrain</i>		0.565	0.043	0.174			0.088	0.087	0.043	
<i>maizesilage</i>		0.39	0.034	0.253	0.008			0.253	0.062	
<i>other</i>		0.75					0.25			
<i>permpasture</i>						1				
<i>potato</i>		0.643		0.071				0.286		
<i>temppasture</i>		0.053	0.008	0.149	0.005		0.008	0.732	0.045	
<i>vegetable</i>		0.314	0.086	0.228		0.029		0.314	0.029	

dairy_pig landuse N	landuse N+1									
	<i>fallow</i>	<i>grain</i>	<i>maizegrain</i>	<i>maizesilage</i>	<i>other</i>	<i>permpasture</i>	<i>potato</i>	<i>raygrassseed</i>	<i>temppasture</i>	<i>vegetable</i>
<i>fallow</i>			1							
<i>grain</i>		0.242	0.29	0.137				0.016	0.234	0.081
<i>maizegrain</i>	0.016	0.64	0.266						0.031	0.047
<i>maizesilage</i>		0.732		0.073					0.195	
<i>other</i>		0.75					0.25			
<i>permpasture</i>						1				
<i>potato</i>		0.643		0.071					0.286	
<i>raygrassseed</i>			0.333					0.667		
<i>temppasture</i>	0.014	0.107	0.057	0.079					0.664	0.079
<i>vegetable</i>		0.285	0.057	0.229					0.2	0.229

pig		landuse N+1								
landuse N	<i>fallow</i>	<i>grain</i>	<i>maizegrain</i>	<i>maizesilage</i>	<i>other</i>	<i>permpasture</i>	<i>potato</i>	<i>raygrassseed</i>	<i>temppasture</i>	<i>vegetable</i>
<i>fallow</i>	0.667	0.333								
<i>grain</i>	0.007	0.184	0.483	0.045	0.004		0.064	0.011	0.03	0.172
<i>maizegrain</i>	0.02	0.469	0.298	0.012			0.041		0.029	0.131
<i>maizesilage</i>	0.091	0.485	0.091	0.03					0.03	0.273
<i>other</i>		1								
<i>permpasture</i>						1				
<i>potato</i>		0.212	0.636							0.152
<i>raygrassseed</i>		0.333	0.167					0.5		
<i>temppasture</i>	0.016	0.106	0.136	0.045					0.636	0.061
<i>vegetable</i>		0.568	0.148	0.136			0.08			0.068

other		landuse N+1								
landuse N	<i>fallow</i>	<i>grain</i>	<i>maizegrain</i>	<i>maizesilage</i>	<i>other</i>	<i>permpasture</i>	<i>potato</i>	<i>temppasture</i>	<i>vegetable</i>	
<i>fallow</i>	0.273	0.273	0.09	0.182					0.182	
<i>grain</i>	0.01	0.173	0.332	0.163	0.016		0.054	0.054	0.198	
<i>maizegrain</i>	0.006	0.377	0.428	0.013				0.031	0.145	
<i>maizesilage</i>	0.018	0.277	0.027	0.347	0.009		0.018	0.188	0.116	
<i>other</i>			0.6						0.4	
<i>permpasture</i>						1				
<i>potato</i>		0.5	0.071	0.358				0.071		
<i>temppasture</i>	0.007	0.109	0.022	0.101			0.007	0.725	0.029	
<i>vegetable</i>	0.02	0.46	0.19	0.06		0.05	0.01	0.08	0.13	

## **Arbres de décision utilisés par *Stochastree* pour les différents types d'exploitation du bassin versant de Naizin**

Les arbres de décision utilisés par l'outil *Stochastree* ont été construits selon une approche de data-mining pour l'ensemble du bassin versant de Naizin et pour les parcelles appartenant aux différents types d'exploitation identifiés, à partir d'un jeu de données d'apprentissage constitué de l'inventaire des cultures sur la période 1993-1998 et des caractéristiques du bassin versant.

L'arbre de décision dérivé des parcelles de l'ensemble du bassin versant (figure A.II.42) ne peut, faute de place, être présenté de façon complète, en indiquant pour chaque feuille la probabilité de transition de chaque culture. On indique donc ici uniquement la structure de cet arbre, permettant d'appréhender comment les critères pris en compte sont testés.

Les arbres de décision qui concernent les parcelles des différents types d'exploitation sont présentés de façon complète et indiquent de ce fait pour chaque feuille la probabilité de transition de chaque type de culture.



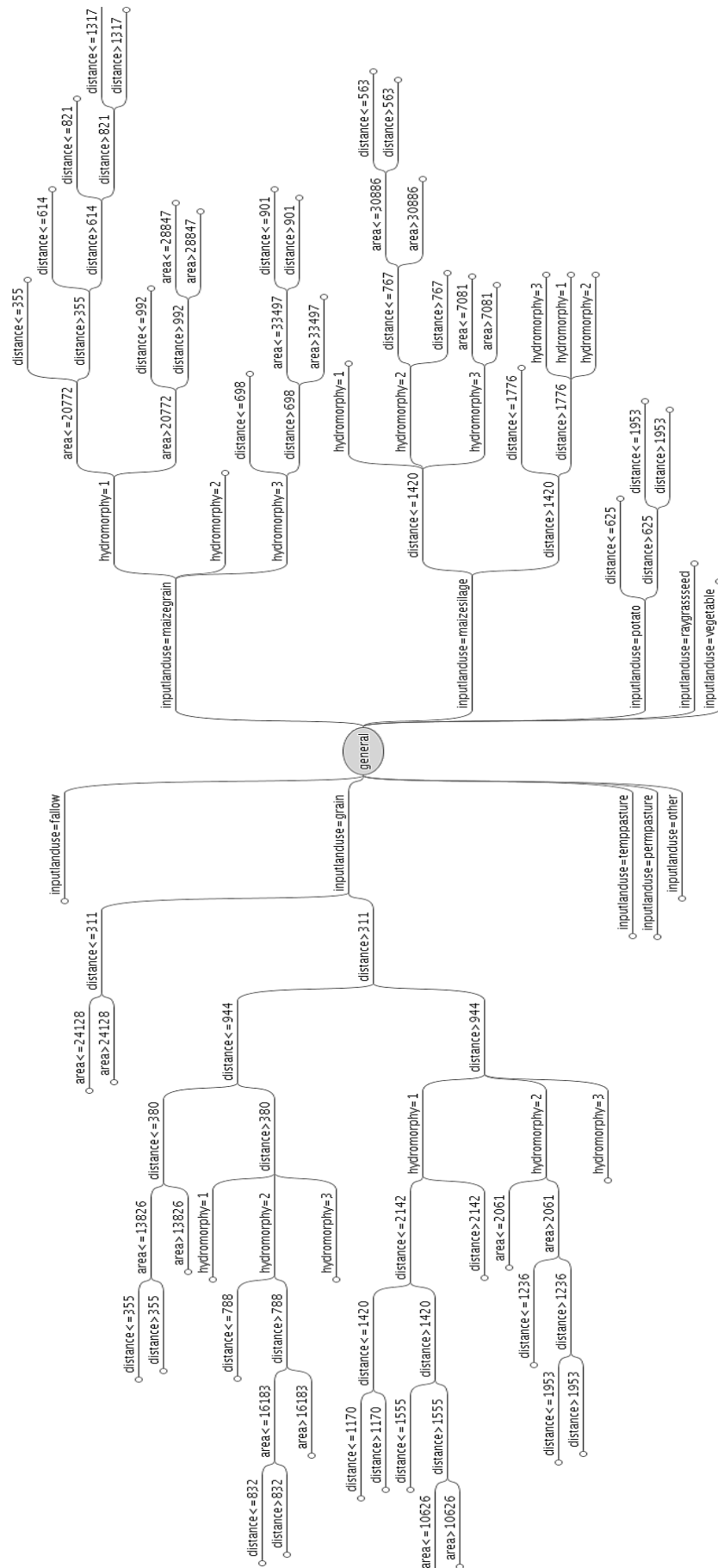


Fig. A.II.42 : Ensemble des parcelles.

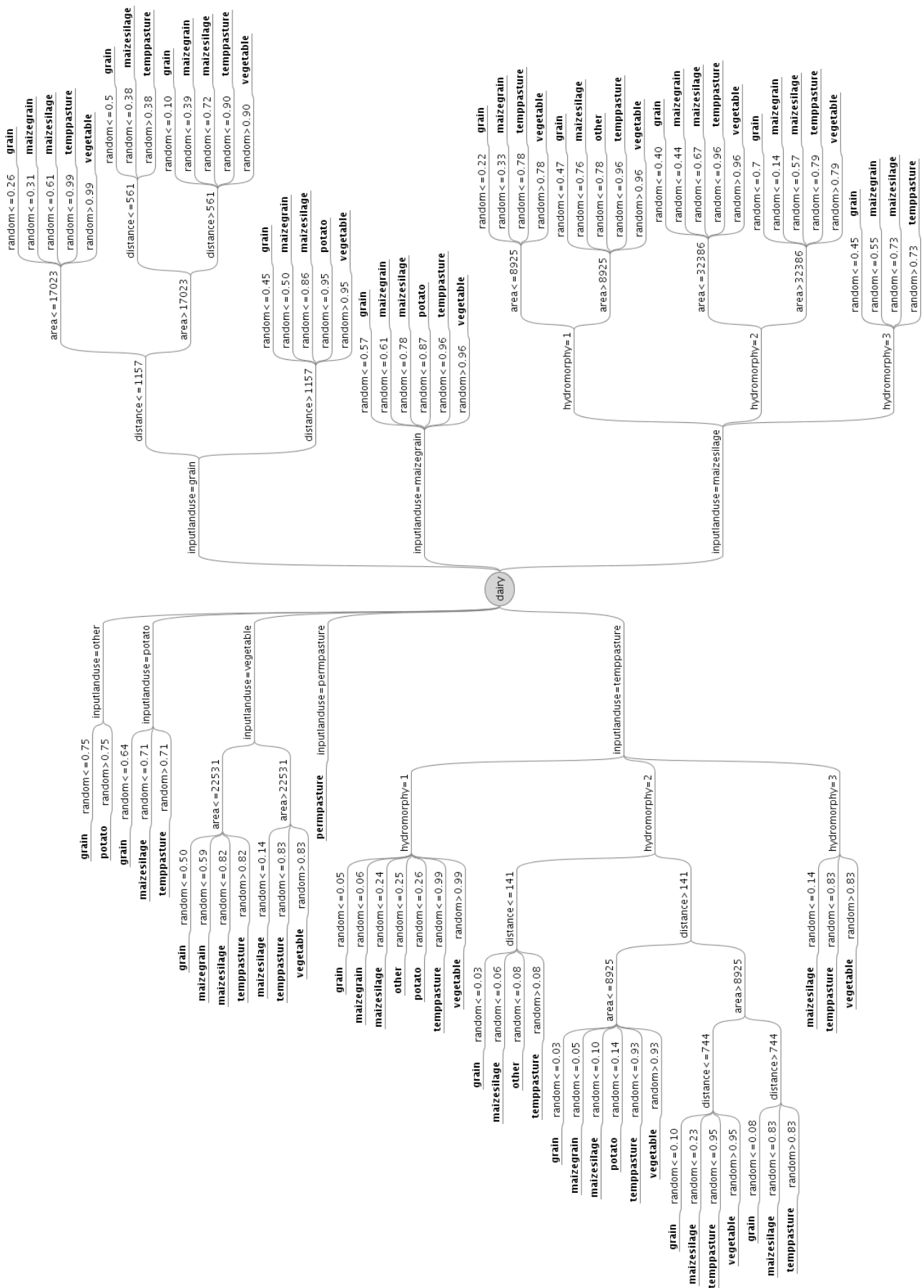


Fig. A.II.43 : Ensemble des parcelles des élevages laitiers.

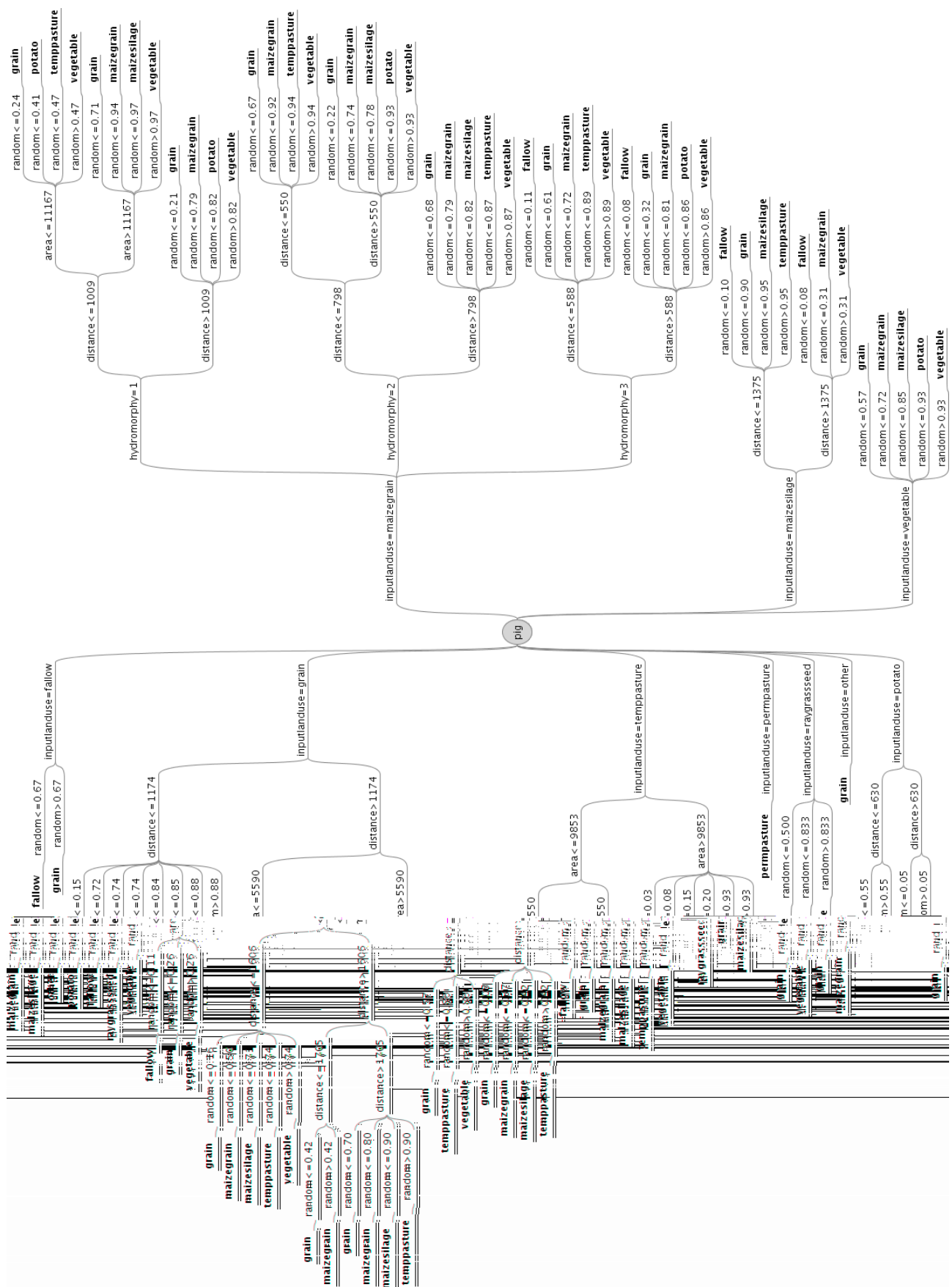


Fig. A.II.44 : Ensemble des parcelles des élevages porcins.

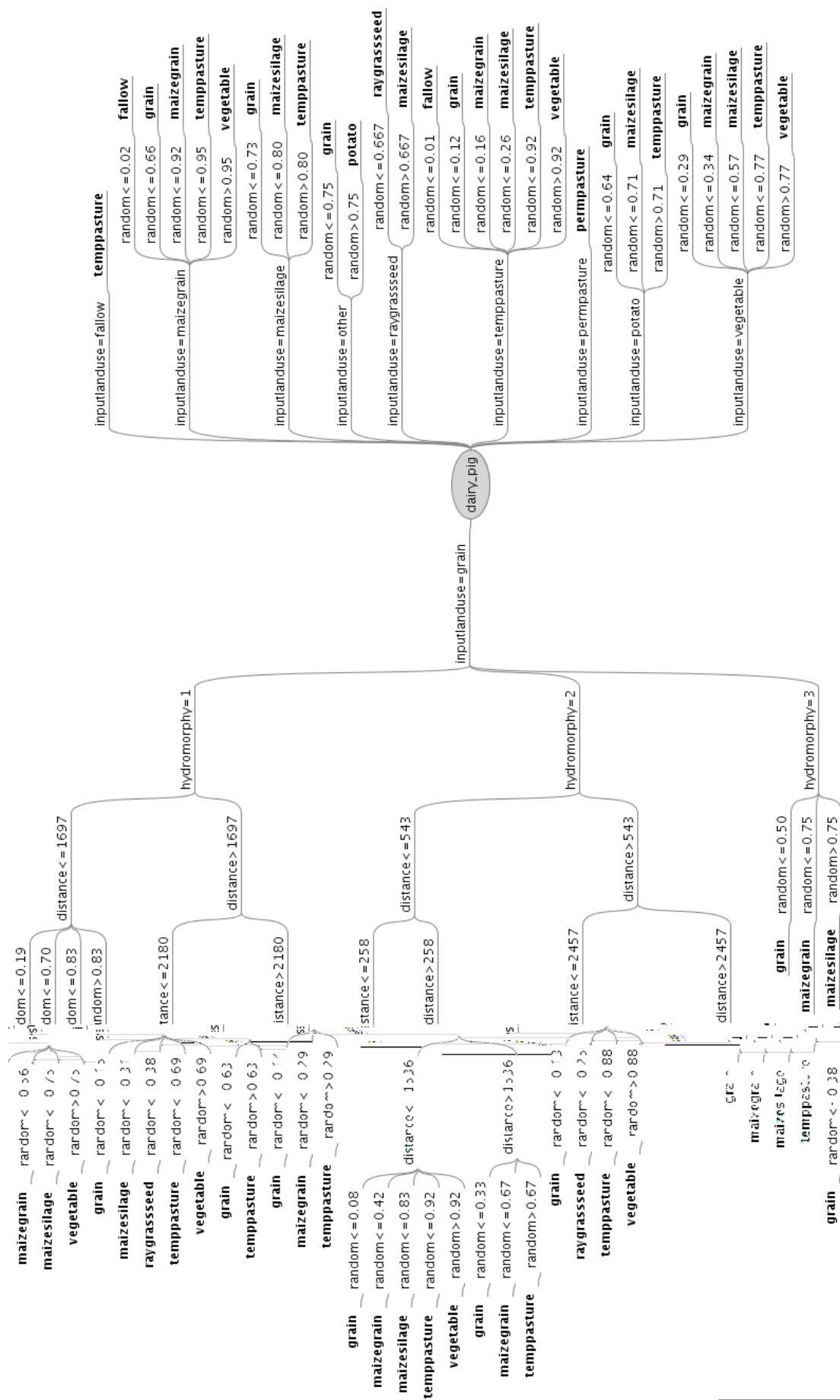


Fig. A.II.45 : Ensemble des parcelles des élevages mixtes lait-porc.

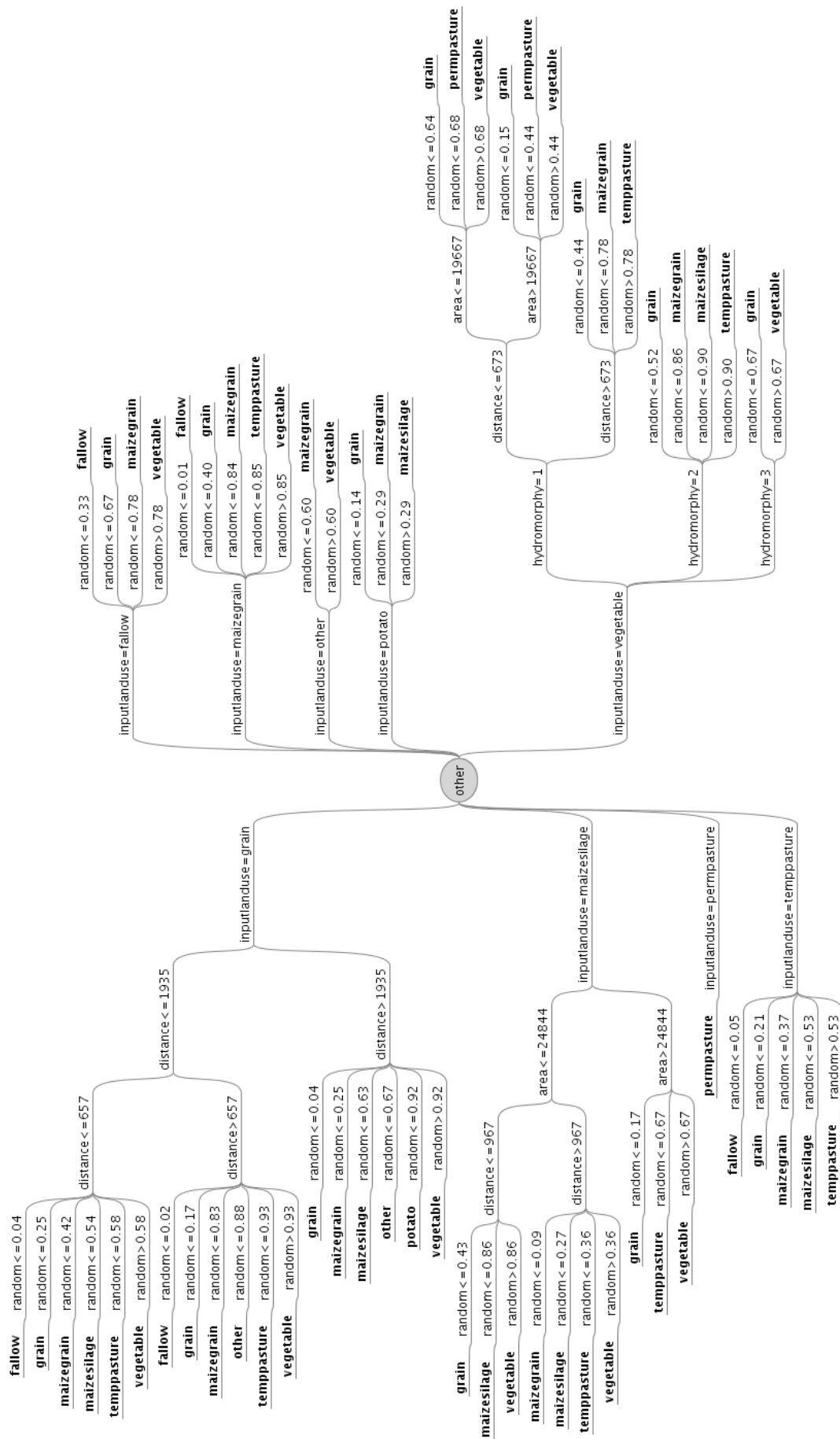


Fig. A.II.46 : Ensemble des parcelles des autres types d'exploitation.

# Arbre de décision utilisé pour simuler l'interculture hivernale

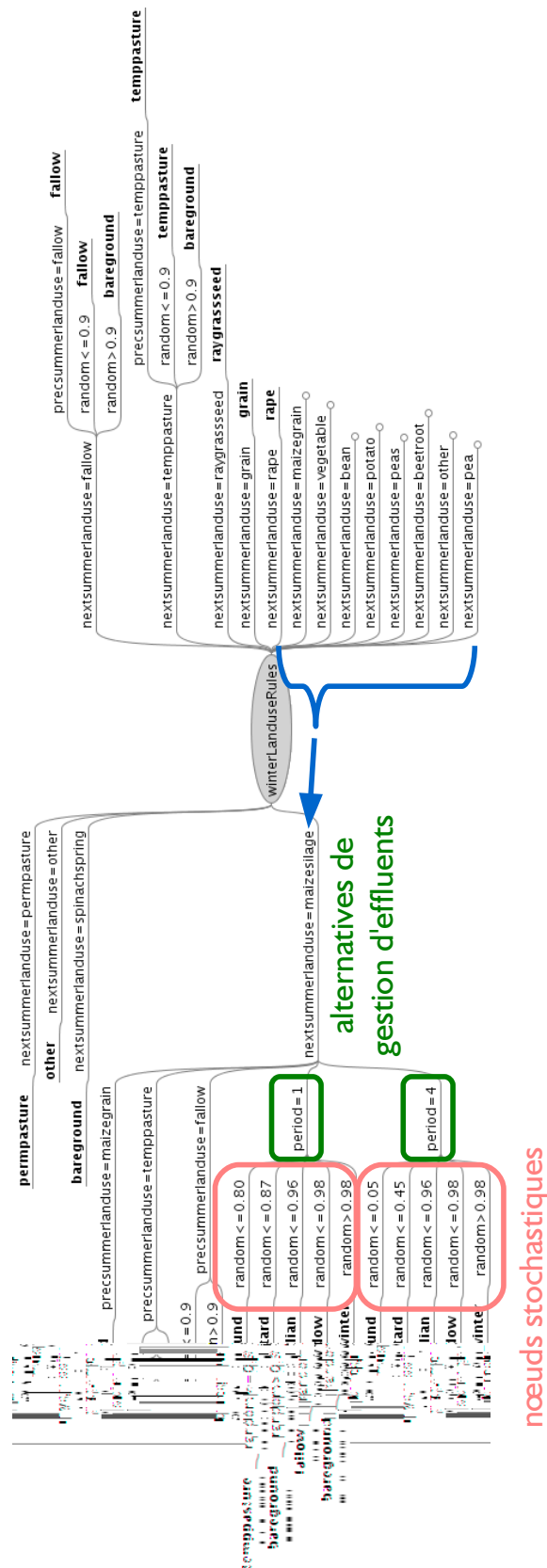


Fig. A.II.47 : Arbre de décision stochastique permettant de déterminer l'occupation du sol hivernale à partir des occupations du sol estivales précédente (precsummerlanduse) et suivante (precsummerlanduse). Les périodes 1 et 2 correspondent aux pratiques intensives et modérées respectivement. Les sous-arbres repliés (maizegrain, grain, vegetable, potato, peas, beetroot, other, pea) ont la même structure que le sous-arbre maizesilage.



## **Annexes III**

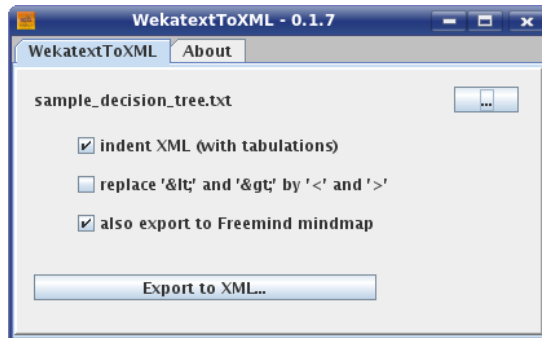
Format et manipulation des arbres de décision



### 1. Exemple de fichier texte produit par l'algorithme J48 de Weka

```
outlook = sunny
|   humidity <= 70: yes
|   humidity > 70: no
outlook = overcast: yes
outlook = rainy
|   windy = TRUE: no
|   windy = FALSE: yes
```

### 2. Import du fichier texte dans WekatextToXML



### 3. Représentation au format XML (gauche) et visualisation graphique sous Freemind (droite)

```
<?xml version="1.0" encoding="UTF-8"?>
<DecisionTree type="sample_decision_tree">
  <Test attribute="outlook" operator="=" value="sunny">
    <Test attribute="humidity" operator="&lt;=" value="70">
      <Output decision="yes"/>
    </Test>
    <Test attribute="humidity" operator=">" value="70">
      <Output decision="no"/>
    </Test>
  </Test>
  <Test attribute="outlook" operator="=" value="overcast">
    <Output decision="yes"/>
  </Test>
  <Test attribute="outlook" operator="=" value="rainy">
    <Test attribute="windy" operator="=" value="TRUE">
      <Output decision="no"/>
    </Test>
    <Test attribute="windy" operator="=" value="FALSE">
      <Output decision="yes"/>
    </Test>
  </Test>
</DecisionTree>
```

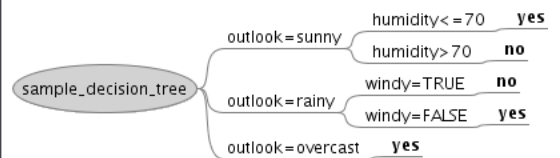


Fig. A.III.48 : Utilisation de WekaTextToXML pour convertir un arbre de décision produit avec Weka (algorithme J48) en fichier XML et en arborescence.

# Annexes III : Format et manipulation des arbres de décision

## Définition du format XML utilisé

Un format XML a été défini pour construire les arbres de décision utilisés dans le cadre de ce projet. Les spécifications en sont simples :

- la racine de l’arbre est la balise **<DecisionTree />**, qui possède un argument *type* qui a servi originellement à stocker le nom du type d’exploitation auquel l’arbre de décision est associé, mais on peut lui réserver d’autres usages ;
- les balises **<Test />** sont utilisées pour représenter les tests réalisés aux nœuds intermédiaires des arbres de décision. Ils sont caractérisés par trois attributs :
  - *attribute* : le nom de l’attribut testé par le nœud ;
  - *operator* : la nature du test ( $<$ ,  $\leq$ ,  $=$ ,  $\geq$ ,  $>$ ) ;
  - *value* : la valeur à laquelle doit être comparée l’attribut ;

À l’instar des arbres produits par l’algorithme C4.5 (Quinlan, 1993), les tests opérés à chaque niveau de l’arbre doivent réaliser une partition complète de l’univers des possibles de l’attribut. Ainsi, si l’attribut est numérique, un nœud de test «  $<X$  » doit être suivi par un test «  $\geq X$  ». Si l’attribut est nominal, alors il doit y avoir autant de tests «  $=$  » que de valeurs possibles.

- les balises **<Output />** représentent les feuilles des arbres, c’est-à-dire les nœuds finaux portant les décisions attendues. Elles ne sont caractérisées que par un attribut, *decision*, qui porte la valeur de la décision que l’arbre doit renvoyer.

La figure A.III.48.3 propose un exemple de fichier XML utilisant ce format ainsi qu’une représentation graphique de l’arborescence des nœuds qui le composent.

## Production et visualisation des arbres

Les arbres de décisions utilisés par *Stochastree* ont été produits à l’aide du logiciel de datamining *Weka*<sup>38</sup>, en particulier l’algorithme J48, qui produit un pseudo-code facile à lire par un humain mais peu exploitable par un programme informatique (figure A.III.48.1). Nous avons donc développé un programme, *WekatextToXML* (figure A.III.48.2), permettant de convertir ce pseudo-code en fichier XML respectant le standard proposé précédemment, et aussi sous la forme de mindmap pour une visualisation en 2D de l’arbre grâce au logiciel *Freemind*<sup>39</sup>. La figure A.III.48 montre le cheminement de la conversion.

38 page d’accueil du site web de *Weka* : <http://www.cs.waikato.ac.nz/~ml/weka/>

39 page d’accueil du site web de *Freemind* : <http://freemind.sourceforge.net/wiki/index.php>

## Méthode principale de manipulation des arbres

La classe *XmlDecisionTree* a été développée afin de manipuler les fichiers XML correspondant au format spécifié précédemment. Le code de sa méthode principale, *getDecision()* est donné ci-dessous. Cette méthode utilise en paramètre une liste de couples de chaînes de caractère, représentant les couples attributs-valeurs que l'arbre doit manipuler pour atteindre une décision (une feuille de l'arbre). L'algorithme de la méthode *getDecision(Map<String, String> inputValues)* de l'objet *XmlDecisionTree* réagit donc de la façon suivante lorsqu'il rencontre un nœud `<Test/>` avec un attribut dont le nom contient "random" :

- si le couple `<"...random...", valeur>` n'existe pas dans le jeu d'informations d'entrée *inputValues*, un tirage aléatoire *X* est fait dans  $[0;1[$  selon une loi uniforme et `<"...random...", X>` est inséré dans *inputValues* ;
- sinon, la valeur de *X* dans `<"...random...", X>` est comparée avec la valeur contenue dans l'arbre.

Le code est commenté afin de comprendre les différentes étapes.

```
public String getDecision(Map<String, String> inputValues) throws ExpectNumericException, AttributeNotFoundInEntriesException,
DecisionLeafNotReachedException {
    Element element = this.xmlTreeDocument.getRootElement();
    String testAtt; // The attribute tested at the node
    String testOp; // The operator of the test; one of {<, >, =, <=, >=, etc.}
    String testVal; // The threshold value of the test
    String testAttType; // The type of the attribute: String, integer, double

    boolean keepParsingXML = true;
    while (keepParsingXML) {
        List<Element> elements = element.getChildren();
        // Throw a DecisionLeafNotReachedException if the reached element contains no child
        if (elements.isEmpty()) {
            keepParsingXML = false; //Useless because of throw statement below
            throw new DecisionLeafNotReachedException(this.xmlTreeDocument.getBaseURI(), inputValues);
        }
        Iterator<Element> itr = elements.iterator();
        boolean keepTestingElements = true;
        while (keepTestingElements && itr.hasNext()) {
            Object o = itr.next();
            if (o instanceof Element) {
                Element localElement = (Element) o;
                // Return the decision value if the localElement name is "Output" (thus, an XML leaf)
                if (localElement.getName().equals("Output")) {
                    keepTestingElements = false; keepParsingXML = false; //Useless because of return statement below
                    return localElement.getAttributeValue("decision");
                }
            }
        }
    }
}
```

```

}
// Perform the test if the localElement name is "Test"
if (localElement.getName().equals("Test")) {
    testAtt = localElement.getAttributeValue("attribute");
    testOp = localElement.getAttributeValue("operator");
    testVal = localElement.getAttributeValue("value");
    testAttType = this.getAttributeType(testVal);
    // If the variable contains "random", adds a <testAtt,random value> to inputValues, if not already present
    if (testAtt.contains("random")) {
        // Test if "random" already exists in inputValues
        if (!inputValues.containsKey(testAtt)) {inputValues.put(testAtt, ""+StrictMath.random());}
    }
    // Throw an AttributeNotFoundInEntriesException if the attribute is not in the inputValues
    if (!inputValues.containsKey(testAtt)) {throw new AttributeNotFoundInEntriesException(testAtt, inputValues);}
    // Perform the test (either a numeric or a String comparison)
    if (testAttType.equals("numeric")) {
        String inputValueString = inputValues.get(testAtt);
        // Throw an ExpectNumericException if the inputValueString cannot be converted as a numeric value
        if (!FormatTools.isNumeric(inputValueString)) {throw new ExpectNumericException(testAtt, inputValueString);}
        keepTestingElements = !this.performNumericTest(inputValueString, testOp, testVal);
    }
    if (testAttType.equals("string")) {
        String inputValueString = inputValues.get(testAtt);
        keepTestingElements = !this.performStringTest(inputValueString, testOp, testVal);
    }
    // Go inner the decision tree
    if (!keepTestingElements) {
        element = localElement;
    }
}
}
}
keepParsingXML = !(keepTestingElements && !itr.hasNext());
}
// Throw a DecisionLeafNotReachedException in case the inputValues were not able to reach a decision leaf of the XML decision tree.
throw new DecisionLeafNotReachedException(this.xmlTreeDocument.getBaseURI(), inputValues);
}

```