



HAL
open science

Atelier "Systèmes d'Information et de Décision pour l'Environnement" (SIDE 2009)

Sandro Bimonte, André Miralles, François Pinet

► **To cite this version:**

Sandro Bimonte, André Miralles, François Pinet. Atelier "Systèmes d'Information et de Décision pour l'Environnement" (SIDE 2009). Association Inforsid. pp.105, 2009. lirmm-00352389v2

HAL Id: lirmm-00352389

<https://hal-lirmm.ccsd.cnrs.fr/lirmm-00352389v2>

Submitted on 22 Jul 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.

Congrès
Inforsid
26 Mai 2009

***Atelier « Systèmes d'Information et
de Décision pour l'Environnement »***

Actes de l'atelier

Edités par :

Sandro Bimonte, TSCF, Cemagref, Clermont Ferrand

André Miralles, UMR TETIS, Cemagref, Montpellier

François Pinet, TSCF, Cemagref, Clermont Ferrand

Atelier « Systèmes d'Information et de Décision pour l'Environnement »

Inforsid - 26 mai 2009

Les Systèmes d'Information et les Systèmes d'Aide à la Décision représentent des solutions de plus en plus performantes pour relever les récents challenges concernant la gestion et l'analyse des données environnementales.

Souvent, le recueil ou l'acquisition des données environnementales restent dans le domaine environnemental un vrai challenge car les techniques mises en œuvre (ex : enquêtes) ou l'instrumentation déployée (ex : satellite) sont lourdes et onéreuses. Or cette activité de recueil et d'acquisition est essentielle car sans information pertinente et de qualité les Systèmes d'Information ou d'Aide à la Décision deviennent rapidement inopérants.

L'objectif de l'atelier est de présenter les dernières avancées dans le domaine des Systèmes d'Information mais aussi de présenter des outils et des méthodes permettant d'acquérir ou d'extraire de l'information d'une part et de mettre en forme cette information pour alimenter un système d'information d'autre part. L'atelier est ouvert aussi bien à la présentation de travaux déjà appliqués au contexte de l'environnement, qu'à des réflexions plus prospectives sur les possibilités d'utilisation d'un produit de la recherche en informatique pour une application environnementale.

La journée d'atelier a été découpée en trois sessions. La première présente des méthodes et des outils permettant de mettre en forme des données sur l'eau, la seconde s'intéresse aux systèmes et méthodes pour la gestion des territoires et la troisième concerne les systèmes d'aide à la décision.

Un premier article présente un système informatique pour la collaboration interdisciplinaire basée sur une théorie sociologique appliquée à des problématiques liées à l'eau.

Un autre travail propose l'intégration d'outils d'analyse multicritères dans un S.I.G couplé à des modèles pour l'évaluation du potentiel aquifère des bassins versants.

D'autres travaux portent sur une méthodologie de traitement d'images pour répondre à des besoins de modélisation hydrologique à différentes échelles.

Les techniques multicritères et de statistiques spatiales sont au centre de deux autres articles, l'un pour la valorisation des continuums écologiques et l'autre pour l'analyse du développement urbain.

La gestion des risques environnementaux est abordée par des recherches sur une infrastructure logicielle d'intégration à base d'agents ; une application est faite à l'évolution de la carte du risque d'incendie de forêt.

Un papier présente une nouvelle architecture d'un système d'information pour la géo-localisation des animaux pour la prévention des risques sanitaires.

Un article traite de l'écoulement des eaux et des polluants à l'échelle du bassin versant et introduit une méthode incrémentale et interactive d'apprentissage.

Un dernier article décrit un outil d'extraction de connaissances pour l'aide à la qualification de l'état des milieux aquatiques.

La qualité des travaux laisse présager une journée d'atelier particulièrement enrichissante. Nous remercions par ailleurs tous les membres du comité de programme pour leur excellent travail ainsi que les auteurs des articles.

Les organisateurs de l'atelier:

Sandro Bimonte, TSCF, Cemagref, Clermont Ferrand
André Miralles, UMR TETIS, Cemagref, Montpellier
François Pinet, TSCF, Cemagref, Clermont Ferrand

Atelier « Systèmes d'Information et de Décision pour l'Environnement »

Inforsid - 26 mai 2009

Organisateurs :

Sandro Bimonte, TSCF, Cemagref, Clermont Ferrand
André Miralles, UMR TETIS, Cemagref, Montpellier
François Pinet, TSCF, Cemagref, Clermont Ferrand

Comité de programme :

Ahmed Lbath, LIG, Grenoble
Catherine Roussey, LIRIS, Lyon
Gil De Sousa, TSCF - Cemagref, Clermont Ferrand
Jean-Christophe Desconnets, IRD, Montpellier
Jean-Pierre Chanet, TSCF - Cemagref, Clermont Ferrand
Maguelonne Teisseire, UMR TETIS – Cemagref, Montpellier

Mathieu Roche, LIRMM, Montpellier
Michel Passouant, Cirad, Montpellier
Michel Schneider, LIMOS, Clermont Ferrand
Sylvie Servigne, LIRIS, Lyon
Thierry Badard, CRG – Université Laval, Québec
Vincent Abt, TSCF - Cemagref, Clermont Ferrand

Programme :

ACCUEIL: 9h00-9h30

Session: Acquisition et mise en forme d'informations sur l'eau

9h30-10h10 *Prendre en compte la dimension sociale dans un projet de développement durable : fondements et utilisation du système Soclab*

C. Sibertin-Blanc, F. Adreit, C. Vautier, P. Roggero

Université des Sciences Sociales-Toulouse 1

Université de Toulouse Le Mirail-Toulouse 2

IRIT/SMAC

LEREPS/CIRESS

10h10-10h50 *Méthode d'évaluation des potentialités aquifères de la nappe alluviale de la Dore (Massif Central)*

D. Mimoun, D. Graillet

Centre SITE, Ecole Nationale Supérieure des Mines de Saint-Etienne

PAUSE: 10h50-11h10

11h10-11h40 *Cartographie de l'occupation du sol pour la modélisation hydrologique spatialisée du cycle de l'eau en zone péri-urbaine*

D. Béal, M. Gagnage, C. Jacqueminet, S. Kermadi, K. Michel, S. Jankowsky, F. Branger, I. Braud

Université de Lyon - CNRS

Cemagref, UR HHLV

Session: Gestion des Territoires

11h40-12h30 *Modélisation pour la valorisation des continuums écologiques départementaux dans l'aménagement du territoire. Application aux Scot de la Loire (42)*

M. Batton-Hubert, M. Bonnevielle, T. Joliveau, F. Paran
Centre SITE - Ecole Nationale Supérieure des Mines de Saint-Etienne
Conservatoire des Espaces Naturels du Languedoc-Roussillon
CRENAM - Université Jean Monnet de Saint-Etienne

REPAS: 12h30-14h00

14h00-14h40 *Pyroxène: un système d'aide à la décision territoriale par intégration de simulateurs spatiaux. Application à l'évolution de la carte du risque d'incendie de forêt.*

E. Maillé, B. Espinasse, S. Fournier
Cemagref, UR EMR
Laboratoire des Sciences de l'Information et des Systèmes - Université Paul Cézanne

14h40-15h10 *Urban versus Rural: the decrease of agricultural areas and the development of urban zones analyzed with spatial statistics*

B. Murgante, G. Las Casas, M. Danese
Laboratory of Urban and Territorial Systems, University of Basilicata (Italie)

PAUSE: 15h10-15h30

Session: Systèmes d'aide à la décision

15h30-16h10 *A New Information System for Tracing Geolocations of Bovine Cattle*

M. Visoli, S. Ternes, J.P. Chanet, F. Pinet, A. Miralles, S. Bernard, G. Souza
Embrapa Agriculture Informatics, Campinas (Brésil)
Cemagref, TSCF
Cemagref, UMR TETIS

16h10-16h50 *Apprentissage incrémental de règles de décision à partir de données d'un simulateur*

M. Cordier, T. Guyet, C. Largouët, V. Masson, H. Suchier
Université Rennes 1 - IRISA
AGROCAMPUS OUEST

16h50-17h30 *Une démarche fondée sur les treillis de Galois pour l'aide à la qualification de l'état des milieux aquatiques*

A. Braud, C. Grac, S. Pristavu, E. Dor, F. Le Ber
LSIIT
LHyGeS-ENGEES
LORIA

BILAN: 17h30-18h00

Prendre en compte la dimension sociale dans un projet de développement durable : fondements et utilisation du système Soclab

Françoise Adreit,** — Pascal Roggero *,** — Christophe
Sibertin-Blanc**,* — Claude Vautier**,****

* *Université des Sciences Sociales-Toulouse 1*
Manufacture des Tabacs,
2, rue du Doyen Gabriel Marty, 31042 Toulouse cedex

** *Université de Toulouse Le Mirail-Toulouse 2*
5 allées Antonio Machado
31058 Toulouse Cedex 9

*** *IRIT/SMAC*

**** *LEREPS/CIRESS*

adreit@univ-tlse2.fr

{ [pascal.roggero](mailto:pascal.roggero@univ-tlse1.fr), [sibertin](mailto:sibertin@univ-tlse1.fr), [claud.vautier](mailto:claud.vautier@univ-tlse1.fr) }@univ-tlse1.fr

RÉSUMÉ. Sur la base de la formalisation d'une théorie sociologique connue, nous proposons de prendre en compte la dimension sociale dans un projet de développement durable. Dans cet article, nous présentons l'originalité et l'intérêt de cette proposition, le système que nous avons élaboré et sa mise en œuvre dans le projet Concert'Eau.

ABSTRACT. On the basis of the formalization of a well-known sociological theory, we propose to take into consideration the social dimension in a sustainable development project. In this paper, we present the originality and the interest of this proposition, the system we have elaborated and its implementation in the Concert'Eau project.

MOTS-CLÉS : système d'information et de décision, système multi-agents, simulation, formalisation, sociologie de l'action organisée, développement durable, pratiques agricoles

KEYWORDS: information and decision system, multi-agents system, simulation, formalization, sociology of the organized action, sustainable development, agricultural practices

1. Introduction

Le constat de la dégradation de la qualité de l'eau et des milieux aquatiques dans les pays d'Europe et, particulièrement, en France conduit l'Union Européenne, à travers la Directive Cadre sur l'Eau (DCE), à envisager les politiques susceptibles de remédier à cet état de fait. Le constat également que les pratiques agricoles en vigueur dans l'agriculture européenne sont une source importante (environ 60 %) de la pollution des eaux de surface en termes de présence de molécules d'azote et de produits phytosanitaires, amène l'Union Européenne à prévoir, dans le cadre de la réforme de la PAC, l'application de mesures agro-environnementales soumettant les acteurs du secteur agricole à des règles d'éco-conditionnalité susceptibles de modifier fortement, voire de bouleverser, les pratiques et l'économie générale du secteur.

Le projet Concert'Eau¹ a été lancé en octobre 2006 pour trois ans, afin de démontrer la possibilité d'amener les acteurs agricoles d'un territoire à proposer de nouvelles pratiques plus en phase avec la préservation de la ressource en eau et à accepter la mise en œuvre de ces pratiques. Le territoire choisi pour cette démonstration est l'amont du bassin versant du Gers au sein du bassin versant Adour-Garonne, dans le sud-ouest de la France, territoire classé en zone vulnérable Nitrate depuis plusieurs années.

Les partenaires du projet, utilisant une méthode d'aide à la décision fondée sur une concertation des acteurs concernés et éclairée par une évaluation scientifique des propositions émanant de ces acteurs, ont exploré les trois dimensions qui fondent les modèles de développement durable :

- sur le plan environnemental, il s'agissait de simuler l'effet des mesures proposées par les acteurs sur la concentration de molécules d'azote et de produits phytosanitaires dans les eaux de surface du Gers et de ses affluents ;
- les économistes étaient chargés de mesurer l'effet potentiel de ces mesures sur le temps de travail des agriculteurs, leur revenu brut d'exploitation et sur le coût direct pour la collectivité (du fait des subventions accordées aux agriculteurs acceptant d'appliquer ces mesures, afin de compenser une éventuelle perte d'exploitation) ;

¹ Il s'agit d'un contrat Life européen porté par Ecobag-ADERA (Association pour le Développement de l'enseignement et de la Recherche auprès des universités, des centres de recherche et des entreprises d'Aquitaine) et l'AEAG (Agence de l'Eau Adour Garonne), ayant pour partenaires : le Cémagref, le Centro de Recursos Ambientales de Navarra, le CNRS, l'Enfa, le Gramip, Infoterra, l'INPT, l'INRA, l'IRH, Silogic et l'Université des sciences sociales de Toulouse. L'Institut de Recherche Informatique de Toulouse (IRIT) est intervenu dans le projet dans le cadre de la modélisation avec les sociologues.

- les sociologues étaient chargés d'estimer l'acceptabilité potentielle de chaque mesure pour l'ensemble du système et pour les agriculteurs (Vautier et al., 2008).

De ce travail multidisciplinaire, nous ne retiendrons ici que la dimension sociologique. L'équipe de sociologues sollicitée disposait pour ce faire d'un système informatique, Soclab, développé dans le cadre d'une collaboration pluridisciplinaire avec des informaticiens, sur la base théorique d'une sociologie connue, la Sociologie de l'Action Organisée (SAO).

Dans un premier temps, nous présentons les principes et l'approche que nous avons adoptés (§2). Nous donnons ensuite les fondements de notre système : la théorie sociologique que nous avons utilisée et le méta-modèle que nous en avons construit, et nous présentons les fonctions du système informatique multi-agents, Soclab, qui a été réalisé sur cette base (§3). Enfin, nous montrons comment nous avons utilisé ce système pour le projet Concert'Eau (§4).

2. Notre proposition : principes et approche

2.1. Prise en compte de la dimension sociale

En théorie, les modèles de développement durable donnent une importance égale aux dimensions environnementale, économique et sociale. En pratique, les dimensions environnementale et économique sont mises en avant tandis que la dimension sociale est peu considérée : on privilégie les aspects biophysiques et le coût économique, pour chaque acteur et pour la collectivité, des dégradations du milieu et de leur compensation par les politiques publiques.

Cependant, l'idée que le sociologue pourrait apporter d'autres éléments d'appréciation apparaît lentement autour de la question de la faisabilité des politiques proposées ou imposées. L'Union Européenne, par exemple, s'interroge depuis plusieurs années sur les raisons de l'inefficacité des mesures qu'elle finance à grand coût. A cette question, le sociologue peut apporter des réponses, notamment en termes d'acceptabilité par les acteurs : si l'acceptabilité sociale des mesures n'est pas acquise, ces mesures ne sont pas acceptées par les acteurs de terrain et par conséquent non appliquées (Vautier et al., 2008).

2.2. Une approche rigoureuse de la dimension sociale

Le manque d'attention accordée à la dimension sociale est dû, pour une grande part, à la nature informelle et subjective des résultats produits par les études sociologiques. Pour remédier à cet état de fait, nous proposons d'utiliser un cadre théorique et méthodologique rigoureux pour évaluer la dimension sociale des mesures proposées (Roggero et al., 2009) :

- nous basons notre travail sur une théorie sociale bien établie, la Sociologie de l'Action Organisée (Crozier, 1963 ; Crozier et al., 1977 ; Friedberg, 1993) ;

- nous avons formalisé cette théorie sous la forme d'un méta-modèle permettant de modéliser de façon rigoureuse des organisations sociales (Gilbert, 1999 ; Phan et al., 2007 ; Roggero et al., 2009 ; Sibertin-Blanc et al., 2006, 2009) ;

- nous avons réalisé un système informatique multi-agents, Soclab, qui permet d'éditer le modèle d'une organisation et d'explorer et simuler le comportement des acteurs (Varenne, 2007 ; Maillard, 2008).

Cette approche a permis d'obtenir des résultats compatibles avec ceux produits par les deux autres disciplines, ce qui a autorisé l'utilisation d'un outil d'intégration *Mulino*² pour obtenir une vision synthétique des résultats des simulations.

2.3. Utilisation du système comme un laboratoire virtuel

S'il ne prétend évidemment pas prendre en compte toute la complexité sociale (Roggero, 2006), Soclab constitue néanmoins un « laboratoire virtuel » pour le sociologue (Barrel, 1979 ; Phan, 2008 ; Varenne, 2003). En effet, grâce à la simulation (Phan et al., 2007), on peut mettre en lumière comment une organisation sociale parvient à l'état régulé constaté empiriquement mais aussi si elle pourrait être régulée autrement en modifiant certains de ses paramètres. Le sociologue peut ainsi tester des scénarios plausibles d'évolution de l'organisation avec une grande facilité. Il génère de la sorte un volume considérable de données significatives qu'il lui serait impossible d'obtenir autrement, même au prix d'importants moyens financiers et humains. SocLab ouvre aussi d'autres pistes d'analyse telle que l'étude des états remarquables de l'organisation et des formes de rationalité des acteurs.

La mise en œuvre de SocLab dans le cadre du projet Concert'Eau a montré que, malgré le caractère réducteur de tout modèle, l'on pouvait ainsi explorer le monde social et aider les acteurs sociaux à délibérer et à effectuer des choix concrets.

3. Le système : fondements et outillage

3.1 Le socle sociologique : la Sociologie de l'Action Organisée

Depuis les années 1970, l'école française de sociologie des organisations a développé un programme de recherche, la Sociologie de l'Action Organisée (SAO), dont la fécondité n'est pas discutée. Ce corpus sociologique est l'un des plus

² Mulino (MULTisectoral INtegrated and Operational decision support system, également nommé mDSS) est un outil informatique d'analyse multicritère permettant d'intégrer des données de nature différente en vue d'une aide à la décision concernant la gestion des ressources naturelles (Mulino, 2004).

enseignés en France tant aux spécialistes qu'aux non-spécialistes, notamment les futurs cadres dirigeants. Il a pour objectif de découvrir le fonctionnement réel d'une organisation, la façon dont elle est régulée, au-delà des règles formelles qui la codifient.

Pour la SAO, les organisations sont des « construits sociaux » actualisés dans et par les relations que les acteurs organisationnels entretiennent entre eux. Ces acteurs sont dotés d'une rationalité limitée et mobilisent leurs ressources pour disposer du pouvoir leur permettant de préserver et/ou d'accroître leur autonomie et leur capacité d'action dans l'organisation. Le pouvoir d'un acteur résulte de la maîtrise d'une ou de plusieurs « zones d'incertitude », c'est-à-dire de ressources nécessaires à l'action d'autrui et dont il maîtrise, au moins partiellement, l'accès. Cette maîtrise lui permet à la fois de fixer, dans une certaine mesure, les « termes de l'échange » dans la relation avec autrui et de rendre son comportement plus ou moins imprévisible. Dès lors, les relations de pouvoir structurent des configurations sociales, relativement stabilisées, qualifiées de « Systèmes d'Action Concrets » (SAC). Un SAC peut être défini comme l'ensemble constitué, dans un contexte organisationnel donné, par les acteurs et leurs alliances, leurs relations et la régulation de ces dernières. Un SAC est donc un contexte d'interaction assez précisément délimité qui structure la coopération d'un ensemble déterminé d'acteurs, de façon certes contraignante mais sans leur ôter toute marge de manœuvre. D'après Friedberg, « tout contexte d'action peut être conceptualisé comme sous-tendu par un SAC » (Friedberg, 1993, p 156), ce qui implique que la finalité d'une recherche s'inscrivant dans ce corpus est bien d'identifier le SAC pour rendre compte (de la régulation) du fonctionnement de l'organisation.

3.2 *La formalisation de la SAO : le méta modèle d'un Système d'Action Concret*³

La figure 1 présente, sous la forme d'un diagramme de classes UML, le méta-modèle d'un SAC, dont les éléments constitutifs sont des *acteurs* et des *relations*, reliés par les associations *contrôler* et *dépendre*.

Un *acteur* qui contrôle une *relation* (ou ressource) en décide de l'*état*, c'est-à-dire de la façon plus ou moins cooperative don't il rend cette ressource mobilisable par les autres. En agissant ainsi sur l'*état* d'une *relation*, l'*acteur* détermine la manière dont sont distribués les *soldes* aux autres *acteurs*, compte tenu de leur fonction d'*effet* qui définit, pour chaque *relation* et chaque *acteur*, le *solde* reçu par ce dernier en fonction de l'*état* de la *relation*. Chaque *acteur* place de plus des *enjeux* sur chaque *relation* dont il *dépend*.

Il en résulte pour chaque *acteur* une certaine *capacité d'action*, agrégation des enjeux et des soldes qu'il reçoit sur les *relations* dont il dépend, et un certain

³ Pour plus de détail sur le méta-modèle, voir (Sibertin-Blanc et al., 2006) et (Sibertin-Blanc et al., 2009).

pouvoir, agrégation des capacités d'action qu'il attribue aux *acteurs* qui dépendent des *relations* qu'il contrôle.

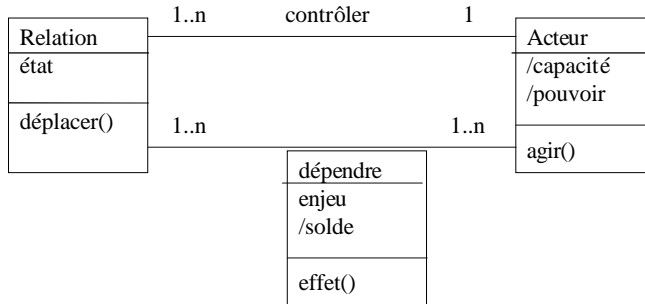


Figure 1. le méta-modèle d'un SAC

3.3 L'environnement Soclab

Nous avons développé un système informatique multi-agent, SocLab (Maillard, 2008) qui permet de décrire le modèle d'une organisation conformément à ce méta-modèle et de l'étudier analytiquement : exploration des configurations remarquables et définition d'indicateurs sociologiques comme la *pertinence* ou la *force* d'une relation ou encore *l'autonomie* ou la *centralité* d'un acteur.

Il permet aussi de simuler le comportement des acteurs (qui modifient l'état des relations qu'ils contrôlent) jusqu'à atteindre une *régulation* du SAC (configuration dans laquelle aucun acteur ne modifie plus l'état des relations qu'il contrôle, se trouvant dans une situation le satisfaisant et ne contestant pas celle des autres). On peut ainsi paramétrer les rationalités des acteurs et explorer des équilibres de Pareto, de Nash ou encore des impossibilités d'équilibre, donc de régulation. On peut donc déterminer de façon quasi expérimentale les conditions selon lesquelles un système peut se réguler et comment.

4. Utilisation du système dans le projet Concert'Eau

4.1 Le système d'action modélisé

L'enquête de terrain a permis d'identifier huit acteurs principaux et huit relations principales (tableau 1). Nous avons choisi d'associer à chaque acteur le contrôle d'une relation unique correspondant à sa maîtrise principale.

Acteurs	Relations
l'agence de l'eau (AE)	Financement des projets
les services déconcentrés de l'Etat : DRAF, DDAF (Direction Régionale/Départementale de l'Agriculture et de la Forêt), mais aussi ONEMA (Office National de l'Eau et des Milieux Aquatiques)	Contrôle réglementaire
le conseil général (CG)	Cofinancement des projets
la chambre d'agriculture (qui en fait officiellement le centre de son activité) (CA)	Conseil aux agriculteurs
les coopératives (coop)	Relais vers les agriculteurs
les instituts techniques et scientifiques	Données (celles qui servent de base aux politiques publiques)
les agriculteurs dits « conventionnels » (avec les coopératives et la chambre d'agriculture)	Lobbying majoritaire
les agriculteurs biologiques et les associations environnementales	Lobbying contestataire

Tableau 1. Liste des acteurs et des relations qu'ils contrôlent

Les acteurs ont défini des enjeux sur les relations dont ils dépendent (tableau 2)⁴.

	AE	DRAF DDAF	CG	CA	Coop.	Instituts	Agriculteurs conventionnels	Agriculteurs biologiques
Financement	4	3	2	1	0	1,5	2,5	0,5
Contrôle réglementaire	2	3	2,5	1	1	2	1	2,5
Cofinancement	2	1,5	4	1	0	1,5	1	1
Conseil	0	0	0	2	2	1	1,5	1
Relais vers les agriculteurs	0	0	0	1	3	1	1	1
Données	2	2,5	1,5	1	1	3	0	0
Lobbying majoritaire	0	0	0	2	3	0	2,5	1
Lobbying contestataire	0	0	0	1	0	0	0,5	3

Tableau 2. Les enjeux placés par les acteurs sur les relations

⁴ Nous avons demandé à chaque acteur d'évaluer les points d'enjeux (entre 0 et 10) qu'il attribue à chaque relation. Selon l'idée de la SAO que tous les acteurs d'une organisation ont le même investissement dans le jeu social, nous avons ramené la somme des points d'enjeux qu'un acteur attribue à l'ensemble des relations à 10. Pour plus de détail, voir (Sibertin-Blanc et al., 2009).

4.2 Les résultats tirés de Soclab

Soclab calcule la capacité d'action et le pouvoir de chaque acteur, quand le SAC est régulé (tableau 3)⁵.

Pour donner rapidement un exemple d'interprétation de ces données, on peut faire remarquer que les acteurs qui ont la plus grande capacité d'action sont les coopératives, la chambre d'agriculture et les agriculteurs conventionnels, tandis que ceux qui en ont le moins sont les agriculteurs biologiques. Cette situation n'est guère paradoxale dans un système qui vit sur un certain *statu quo* constitué d'une vision dominante productiviste de l'agriculture associée à l'habitude du subventionnement.

D'autre part, dans le cas de Concert'Eau, il a été intéressant de faire apparaître un pouvoir que nous avons appelé coopératif (tableau 3), qui ne prend en compte que les capacités d'action *positives* qu'un acteur prodigue *aux autres*. L'étude des écarts entre le pouvoir tel qu'il a été défini dans le méta-modèle et le pouvoir coopératif montre que certains acteurs sont très coopératifs (l'agence de l'eau, le conseil général ou les instituts techniques) tandis que d'autres le sont peu (la chambre d'agriculture). Les agriculteurs conventionnels, comme les biologiques, ne le sont pas, mais pour des raisons différentes : les premiers, en alliance avec les coopératives et la chambre d'agriculture participent d'un « oligopole agricole » qui dispose d'un pouvoir dominant (si l'on fait la somme des pouvoirs des trois acteurs) et sont moins soucieux de coopération que d'exercer de fortes pressions sur l'organisation ; les seconds parce qu'ils considèrent qu'ils sont isolés et qu'aucun autre acteur ne peut (ou ne veut) épouser un tant soit peu sa cause.

Enfin, on constate qu'en dehors de l'agence de l'eau (AE), l'acteur ayant le plus de pouvoir est constitué des instituts scientifiques et techniques. Ce paradoxe apparent est expliqué par le fait que les données concernant le rapport eau-agriculture sont l'objet de batailles serrées et récurrentes entre les acteurs

Sur la base de ces résultats, il est possible de déterminer quels sont les acteurs qui auront des raisons et la possibilité de s'opposer ou au contraire de favoriser l'application de chacune des mesures proposées (Vautier et al. 2009).

	AE	DRAF DDAF	CG	CA	coop	instituts	agriculteurs conventionnels	agriculteurs biologiques
Capacité	68,4	50,7	61,4	71,4	78,1	52,0	72,0	38,0
Pouvoir	146,6	41,5	74,2	25,0	65,1	141,8	101,7	56,5
Pouvoir coopératif	144,2	27,5	74,2	7,9	40,0	141,8	55,4	0,9

Tableau 3. Capacité d'action et pouvoirs des acteurs à convergence des simulations

⁵ La valeur d'une capacité d'action ou d'un pouvoir n'a de sens que comparativement aux autres ; ainsi, de la première ligne, on peut établir que ce sont les coopératives qui ont la plus grande capacité d'action et que cette capacité d'action est le double de celle des agriculteurs biologiques.

4.3 Evaluation de l'acceptabilité sociale des mesures proposées

Partant des satisfactions de la situation actuelle (tableau 3), nous avons effectué une seconde enquête de terrain pour évaluer, pour chaque acteur, l'acceptabilité de chacune des mesures proposées par rapport à la situation actuelle. Chaque acteur a donné une note à chaque mesure⁶. Nous avons alors pondéré les satisfactions obtenues par simulation (tableau 3) avec ces notes pour obtenir la satisfaction de chaque acteur pour chaque mesure. De plus, la diminution de la satisfaction d'un acteur possédant beaucoup ou peu de pouvoir ayant des conséquences différentes, nous avons pondéré les satisfactions par les pouvoirs. Nous avons ainsi obtenu ce que nous appelons la faisabilité de chaque mesure pour chaque acteur. Enfin, pour chaque mesure, nous avons fait la somme des faisabilités de tous les acteurs pour obtenir la faisabilité du point de vue du système et la somme des faisabilités pour les agriculteurs pour obtenir la faisabilité du point de vue des agriculteurs. Le tableau 4 présente ces résultats pour 7 mesures proposées (M1 à M7).

On peut alors comparer la faisabilité de chaque mesure M_i par rapport à la valeur actuelle (colonne 1). Par exemple, M1 et M2 ne sont pas favorables aux agriculteurs (60,3 est plus petit que 90,5).

	Actuel	M1	M2	M3	M4	M5	M6	M7
système	457	468.1	476.2	494	511.4	483	459	501.1
agriculteurs	90.5	60.3	60.3	90.5	98.2	96.3	88.5	114.6

Tableau 4. Faisabilité (comparée à l'état actuel) des mesures du point de vue du système et des agriculteurs

5. Conclusion

L'utilisation de Soclab dans une telle étude apporte une information généralement absente des études servant à décider d'un choix de politique publique. Elle permet d'informer les choix des décideurs en explicitant quels seraient les changements susceptibles, au-delà de leurs effets environnementaux et économiques, de provoquer un blocage de la part de certains acteurs ou au contraire d'être suffisamment acceptables par les acteurs pour être introduits avec des chances de succès, c'est-à-dire sans rejet massif de la part des acteurs les plus dotés en ressources et capables de bloquer le système ou de détourner leur application à leur profit exclusif. Face aux questions de l'Union Européenne sur les raisons d'une faible productivité des fonds dépensés pour modifier le système, Soclab apparaît comme un outil heuristique pour proposer des réponses fortement argumentées.

⁶ Chaque note est comprise entre 0 et 10, 5 représentant la valeur neutre, les notes inférieures (supérieures) à 5 traduisant une plus ou moins grande insatisfaction (satisfaction).

Bibliographie

- Barrel Y., *Le paradoxe et le système. Essai sur l'imaginaire social*, PUG, Grenoble, 2^e éd., 1979.
- Crozier M., *Le phénomène bureaucratique. Essai sur les tendances bureaucratiques des systèmes d'organisation modernes et sur leurs relations en France avec le système social et culturel*, Seuil, 1963.
- Crozier M., Friedberg E., *L'acteur et le système : Les contraintes de l'action collective*, Paris, Seuil, 1977.
- Friedberg E., *Le pouvoir et la règle. Dynamiques de l'action organisée*, Paris, Seuil, 1993.
- Gilbert N., « Simulation : A New Way of Doing Social Science », *American Behavioural Scientist*, vol. 40, n°10, 1999, p.1485-1487.
- Maillard M., Formalisation multi-Agents de la Sociologie de l'Action Organisée, Thèse de doctorat en informatique de l'Université de Toulouse, décembre 2008.
- Mulino, MULTisectoral INtegrated and Operational decision support system, or mDSS, 2004, voir : <http://www.netsymod.eu/mdss/> et siti.feem.it/mulino/
- Phan D., « De l'exploration conceptuelle à l'expérience concrète : la modélisation et la simulation multi-agents comme un autre moyen d'expérimenter en sciences économiques et sociales », *15ièmes journées de Rochebrune*, janvier 2008. Voir : <http://www.gemas.fr/dphan/paper/phanRochebrune2008.pdf>.
- Phan D, Amblard F (Eds.), *Agent-based Modelling and Simulation in the Social and Human Sciences*, Oxford, The Bardwell Press, 2007.
- Roggero, P., De la complexité en sociologie : évolutions théoriques, développements méthodologiques et épreuves empiriques d'un projet sociologique, Mémoire d'habilitation à diriger des recherches, Université de Toulouse 1, 2006.
- Roggero P., Sibertin-Blanc C., « Formalisation, modélisation et simulation orientée agents des systèmes d'action concrets : approche complexe de la sociologie de l'action organisée », *Nouvelles Perspectives en Sciences Sociales*, vol. 4, n°1, 2009.
- Sibertin-Blanc C., Roggero P., Adreit F., Chapron P., Maillard M., Sandri S., Vautier C., « Un méta-modèle des systèmes d'action concrets », Rencontres interdisciplinaires sur les systèmes complexes naturels et artificiels – *16ièmes journées de Rochebrune*, janvier 2009, Denis Phan (Eds.), ENST, 2009 (à paraître)
- Sibertin-Blanc C, Amblard F, Maillard M, « A coordination framework based on the Sociology of Organized Action », *Coordination, Organizations, Institutions and Norms in Multi-Agent Systems*, O. Boissier, J. Padget, V. Dignum, G. Lindemann (Eds.), Springer, 2006, p.3-17, LNCS 3913.
- Varenne F., *Du modèle à la simulation informatique*, Paris, Vrin, Mathesis coll., 2007, 256 p.
- Varenne F., « La simulation conçue comme une expérience concrète », Le statut épistémologique de la simulation, *10ièmes journées de Rochebrune*, Paris, ENSET ed., 2003, pp. 299-313.
- Vautier C., Roggero P., « Concert'eau : contribution à la modélisation SMA à une gestion participative de la gestion de l'eau », *18th international congress of french-speaking Sociologists*, Istanbul, 7-11th July 2008.
- Vautier C., Roggero P., Adreit F., Sibertin-Blanc C., « Evaluation by simulation of the social acceptability of agricultural policies for water quality », *International ICSC symposium on Information Technologies on Environmental Engineering*, Thessaloniki, Greece, mai 2009 (à paraître).

Méthode d'évaluation des potentialités aquifères de la nappe alluviale de la Dore (Massif Central)

Mimoun Djamel — Graillet Didier

Centre SITE, Ecole Nationale Supérieure des Mines de Saint-Etienne
158, cours Fauriel, 42023 Saint-Etienne Cedex 2

mimoun@emse.fr

graillet@emse.fr

RÉSUMÉ. L'objectif de cette recherche est de proposer une méthodologie visant à évaluer le potentiel aquifère des têtes de bassin versant du Massif Central dans un contexte de modifications climatiques afin de gérer au mieux la ressource en eau et ses différents usages. La méthode articule des outils d'analyse multicritères spatiale intégrés dans un S.I.G et couplés à des modèles mathématiques d'écoulement souterrain. La méthodologie a permis d'identifier 5 zones à fort potentiel aquifère et de caractériser leur vulnérabilité aux pollutions.

ABSTRACT. The goal of this research is to propose a methodology for evaluate the aquifers potential of the heads of area catchment of the Massif Central in a context of climatic modifications in order to as well as possible manage the water resource and its various uses.

MOTS-CLÉS : SIG, modélisation hydrodynamique, transfert de polluant, optimization des debits prélevés

KEYWORDS: groundwater flow, contaminant transport, pumping well optimization, multicriteria analysis, GIS

1. Introduction

Sans que l'on s'en rende compte, l'eau devient un élément clé de l'économie : elle est indispensable au secteur agricole, au secteur industriel et à l'adduction en eau potable des communes. Pourtant, les ressources en eau sur les têtes de bassin versant (zones de sources) sont relativement peu connues.

L'objectif de cette recherche est de proposer une méthodologie visant à évaluer le potentiel aquifère des têtes de bassin versant du Massif Central dans un contexte de changement climatique afin de gérer au mieux la ressource en eau et ses différents usages. La méthode articule des outils d'analyse multicritères spatiale intégrés dans un S.I.G et couplés à des modèles mathématiques d'écoulement souterrain.

Le cas d'étude présenté concerne la Dore, principal affluent rive droite de la rivière Allier, elle-même affluent du fleuve Loire, qui se situe dans le grand bassin hydrographique Loire Bretagne. Sa situation géographique la plaçant parmi les têtes de bassin versant, lui donne une importance stratégique pour la préservation des ressources en eau du bassin de la Loire. Le SDAGE Loire-Bretagne qui répond lui-même à la Directive cadre sur l'Eau a d'ailleurs inscrit le Schéma d'Aménagement et de Gestion des Eaux (SAGE) de la Dore comme prioritaire en termes de reconquête de la qualité de l'eau et des milieux aquatiques.

Après une description du contexte, nous présentons dans cet article les différents aspects de la méthode associant SIG, analyse multicritères et modélisation hydrodynamique. Enfin, quelques résultats sont présentés et discutés.

2. La Dore : un constat alarmant

Le bassin versant de la Dore d'une superficie de 1716 km² est localisé dans la moyenne montagne granitique du Massif Central oriental. De forme très allongée, il s'étire sur 88 km du sud au nord alors que sa largeur varie de 18 à 30 km. Deux grands ensembles peuvent être distingués : un secteur montagneux dont les plus hauts sommets culminent à plus de 1500 m occupé par les zones humides et les tourbières et la basse plaine alluviale façonnée par la rivière qui constitue le prolongement oriental de la Grande Limagne Clermontoise (figure 1)

Cette grande variété de milieux est le siège d'une richesse écologique remarquable représentée par une longue liste d'habitats et d'espèces rares ou menacées de disparition : moule perlière, écrevisse à pattes blanches, loutre...

De part sa situation hydrophysique, le fort potentiel hydraulique de la Dore a suscité le développement d'une activité artisanale et industrielle couplée jusque dans les années cinquante à une forte activité agricole. Ce développement économique, suivi du développement de plusieurs villes, a entraîné la dégradation de la qualité des eaux et des habitats du bassin versant de la Dore [CON 08].

Les résultats d'analyses de la Dore et de ses principaux affluents indiquent aujourd'hui une mauvaise qualité pour le paramètre micropolluants (Chrome,

Nickel, Arsenic) si bien que la Dore est classée au titre de la Directive Cadre sur l'Eau en risque de non atteinte du bon état écologique des eaux.

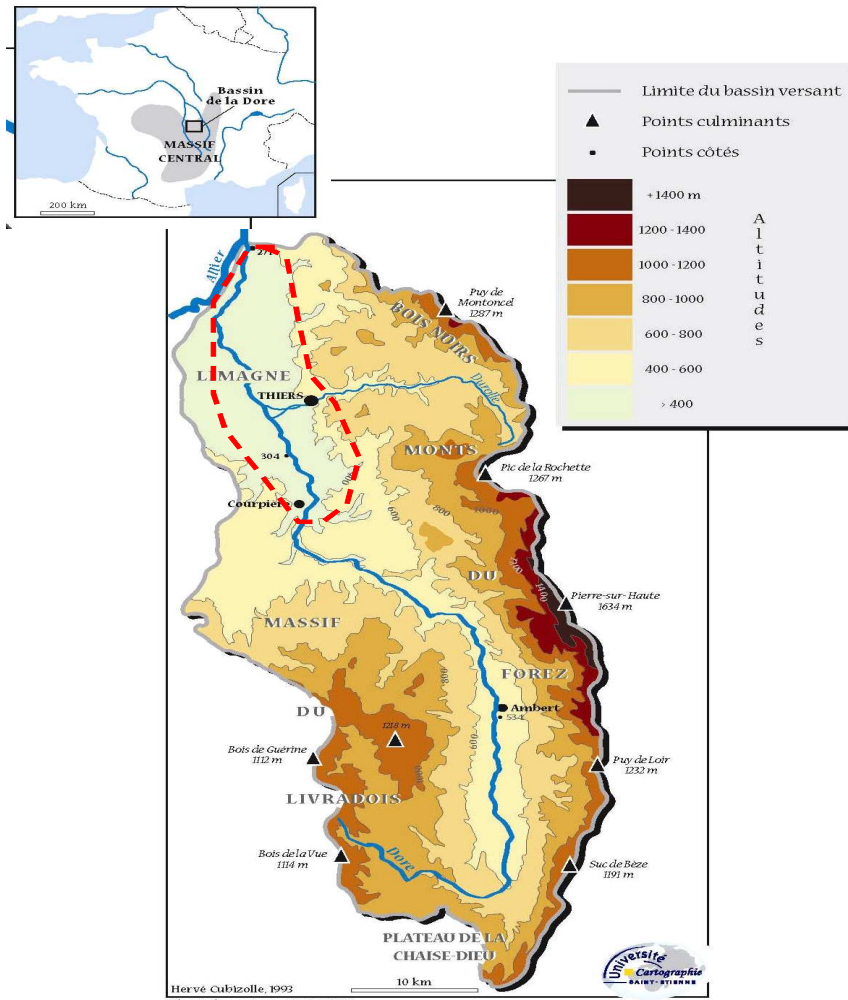


Figure 1 : Localisation du secteur à étudier. Les traits pointillés délimitent la plaine alluviale de la basse vallée de la Dore.

Durant le quaternaire, la Dore a balayé une grande partie de sa plaine alluviale, laissant de nombreux paléochenaux (anciens tracés) parfois encore visibles dans le paysage. Aujourd'hui, la rivière est corsetée par une multitude d'enrochements et de protections de berges (12% du linéaire de berge) qui bloquent tout déplacement

latéral de la rivière et diminuent fortement les échanges hydriques entre la nappe et la rivière [CUB 07].

Comme toutes les plaines alluviales, le lit majeur de la Dore a fait l'objet d'importantes extractions de granulats dans les années 1952-1990. Elles sont aujourd'hui quasiment achevées mais ces aménagements limitent fortement la capacité naturelle d'ajustement du lit de la rivière et posent des problèmes de qualité de l'eau (zone stagnantes marécageuses)

3. Un enjeu majeur : la ressource en eau

3.1. Des prélèvements importants sur les têtes de bassin versants

Actuellement la majorité des débits prélevés pour l'AEP (80% des 575 captages recensés par la DDASS) sur le bassin versant de la Dore (14.3 Millions de m³/an) ont lieu en tête de bassin versant en rive gauche de la Dore dans le massif du Livradois. Il s'agit de captages de sources liées à des écoulements souterrains peu profonds dans les zones de contact entre les anciennes terrasses alluviales et les formations granitiques. Pendant les périodes longues de sécheresse (2003), les sources se tarissent. La faible capacité de stockage des aquifères du socle explique des étiages naturellement très faibles [MIM 08]. Cependant, les pratiques comme les captages cumulés de sources, le drainage des zones humides ou leur boisement ont contribué à aggraver ces étiages. Les conséquences sont importantes : altération de la qualité des eaux, atteintes au milieu naturel, non satisfaction de l'approvisionnement en eau potable, multiplication des situations de gestion de crise au cours des dernières années...

Au niveau de la plaine alluviale de la Dore, l'utilisation de la ressource en eau se concentre au niveau de trois zones de captage vers Thiers, Dorat et Ris. Des activités non autorisées telles que les dépôts de déchets et extractions sauvages constituent des sources de pollution directe de la nappe alluviale. Les milieux anthropiques représentent des sources de pollution diffuse, pour lesquelles la capacité d'épuration par le corridor fluvial est réduite à néant au niveau des secteurs chenalisés de la rivière.

3.2. Un nécessaire transfert d'exploitation de la ressource en eau vers la plaine alluviale

Compte tenu de la très forte vulnérabilité de la ressource en eau sur les têtes de bassin versant et de la relative bonne préservation de la plaine alluviale de la Dore qui constitue la principale ressource aquifère, l'enjeu majeur consiste à maintenir en quantité et en qualité cette ressource en eau pour les générations futures.

En outre, les membres de la commission locale de l'eau (CLE) du SAGE Dore souhaitent savoir s'il est possible de transférer les prélèvements d'eau des têtes de bassin versant vers la plaine alluviale sans altérer les zones à fort intérêt écologique.

Il s'agit donc de définir, dans le cadre du SAGE Dore pour le Parc Naturel Régional du Livradois-Forez, les secteurs aquifères les plus favorables ainsi que les débits potentiellement mobilisables pour la ressource en eau.

4. Principe de la méthode

Cette partie présente les différents aspects de la méthode. La méthode choisie articule d'une part des données environnementales disponible sous forme cartographique et d'autre part des résultats issus d'un modèle mathématique d'écoulement souterrain associé à un module de transport de polluant et à un module d'optimisation des débits prélevés dans les puits (figure 2).

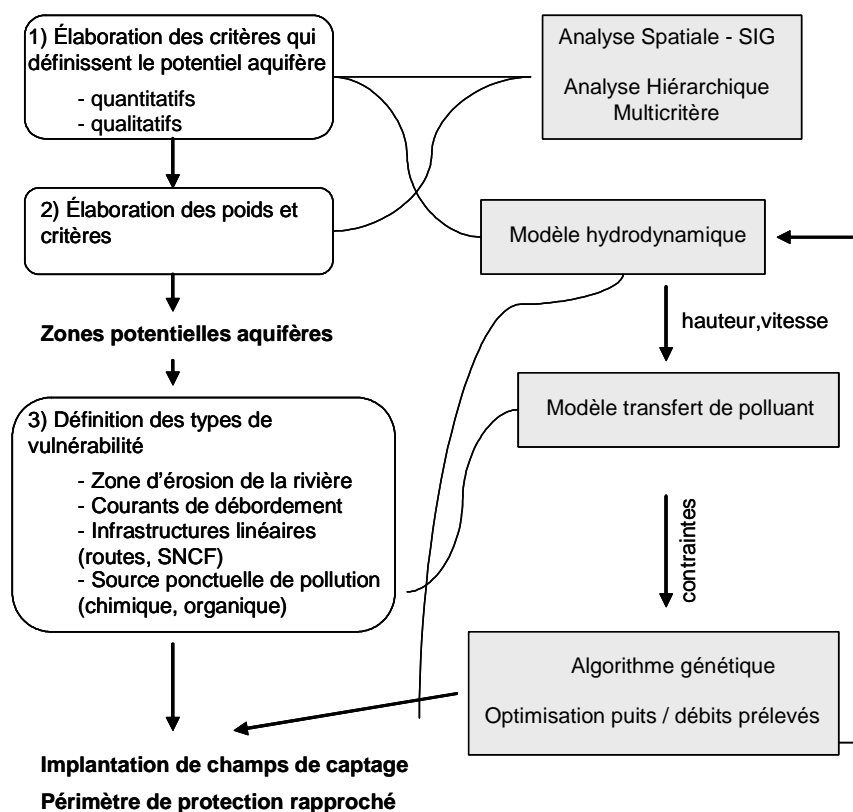


Figure 2. Démarche séquentielle du modèle conceptuel fondé sur l'analyse hiérarchique multicritère (MHM) et sur la combinaison de modèles pour le choix

d'implantation de nouveaux champs de captage. Les flèches indiquent la fourniture de données permettant la production des résultats.

Dans un premier temps, 5 sources de données⁷ de précision différentes ont permis d'une part de délimiter la géométrie du réservoir aquifère physique du réservoir et d'autre part de représenter les écoulements au sein de la nappe alluviale en période de basses et hautes eaux. La modélisation hydrodynamique a été réalisée avec MODFLOW Premium développé par Waterloo Hydrogeologic Incorporated. Il s'agit de l'un des systèmes de modélisation le plus couramment employé pour la modélisation d'aquifères alluviaux relativement homogènes [KIN 86].

Les zones de bon potentiel aquifère sont définies d'après des critères qualitatifs et des critères quantitatifs regroupés en 4 sous critères validés par les acteurs du SAGE. Par exemple, les valeurs de transmissivité sont issues du modèle hydrodynamique. Les zones favorables (forte transmissivité) concernent les zones où l'épaisseur mouillée est importante et la perméabilité forte.

Pour chaque sous-critère, une note de 0 (défavorable) à 5 (très favorable) est ensuite affectée aux différentes valeurs de chaque paramètre (tableau 1). Cette échelle à 5 niveaux de notation a été adoptée par rapport à la qualité des données permettant d'évaluer ces sous-critères.

La carte des zones de bon potentiel est finalement obtenue par agrégation des critères et couplage SIG-AMC en utilisant la méthode Analytic Hierarchy Process développée par Saaty (tableau 1)[SAA 84]. Chaque critère a été pondéré à dire d'expert au cours d'une réunion de présentation de l'étude regroupant l'ensemble des acteurs concernés par la gestion de la ressource en eau. Les critères hydrogéologiques et géomorphologiques demeurent prépondérants. Cette approche est simulée en utilisant la fonction « Weighted sum » implémentée dans ArcGIS 9.3 (module Spatial Analyst).

⁷ Modèle Numérique d'altitude résolution 50 m, Cartes géologiques BRGM au 1/50 000 Feuilles de Thiers et Maringues, Banque du sous sol (BSS) du BRGM, Etude Répartition des ressources en eau et en granulats pour servir à l'aménagement de la plaine alluviale de la Dore (Etude CETE 63), Etude Allier –bec de Dore (secteur Puy-Guillaume)

Thèmes	Sous-critère	Classe	Note	Poids		
Zones potentielles aquifères	- critères qualitatifs	- Transmissivité de l'aquifère	0.010-0.018 m ² /s	1	0.4	
			0.018-0.036 m ² /s	3		
			> 0.036 m ² /s	5		
	- Paléochenaux	Présence		5	0.3	
		Absence		0		
	- critères quantitatifs	- Occupation du sol	Bancs d'alluvions		0	0.2
			Bâtis, routes et abords		0	
			Rivière Dore		0	
			Gravière ou bras morts		0	
			Cultures		1	
Terrains nus				1		
Prairies artificielles				2		
Peupleraies				3		
Forêts de résineux				4		
Forêts				5		
Prairies naturelles		5				
- réseau hydrographique	Gravière (>20 m)		5	0.1		
	Rivière (>15 m)		5			
	Affluents (>10 m)		5			

Tableau 1. Description des critères organisés selon une arborescence en arbre de hiérarchisation et des échelles de mesure utilisées pour l'évaluation des zones aquifères les plus favorables. La pondération des critères associés est donnée dans la colonne poids.

5 Résultats

Les 5 zones identifiées sur l'ensemble du secteur d'étude ont ensuite été caractérisées selon leur type de vulnérabilité en utilisant notamment les inventaires BASIAS (Base des Anciens Sites Industriels et Activités de Service) et BASOL développés par le BRGM et la BD Carthage (figure 3).

La modélisation des transports de substances polluantes en solution en milieu poreux dans la zone non saturée a permis de cartographier la propagation du panache de polluant et de fixer l'isochrone 50 jours (figure 4). Le scénario de pollution retenu concerne une injection accidentelle et instantanée (accidentelle) au niveau des zones identifiées comme potentiellement à risque pour des pollutions chimiques sans adsorption et des pollutions organiques faisant intervenir de la biodégradation.

Au final, après élimination des zones vulnérables c'est à dire celles pour lesquelles le temps de propagation d'une pollution est inférieur à 50j, on obtient les zones où il est possible d'implanter un nouveau champ de captage. Le temps de propagation de 50 jours a été choisi en fonction des délais nécessaires à la mise en place d'un plan de secours (interconnexion de réseaux d'eau potable ou de captages, mise en place d'une barrière hydraulique) [LAL 88].

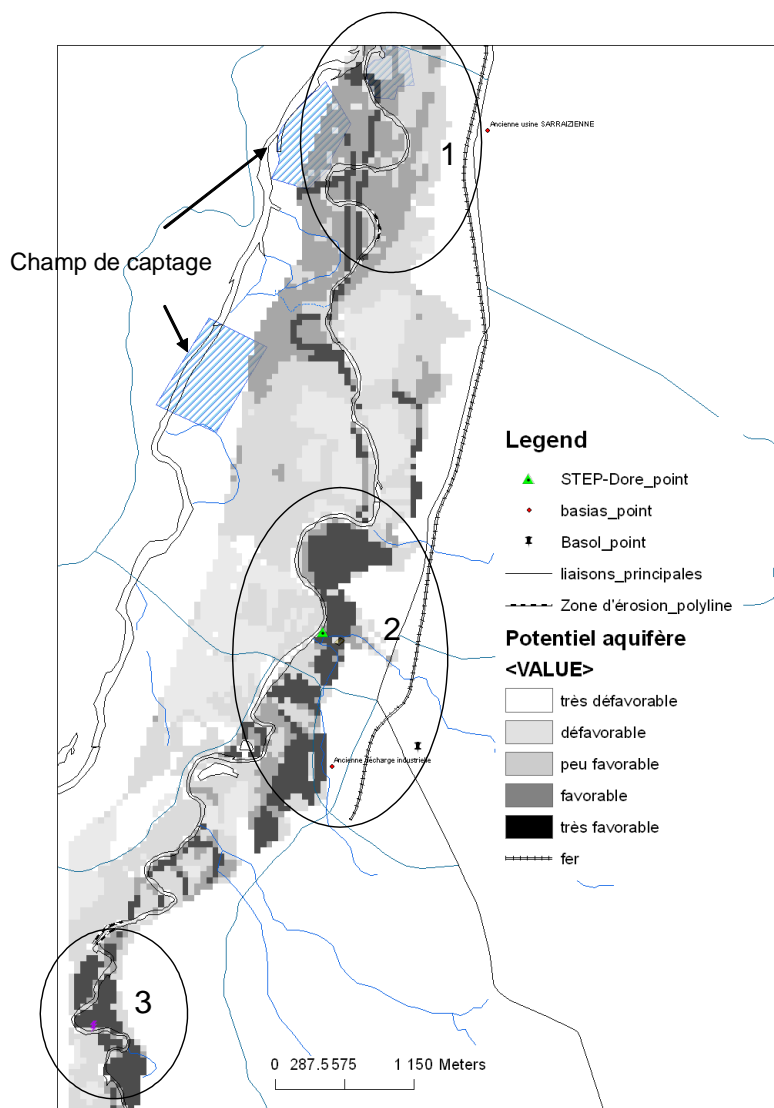


Figure 3. Les 3 zones *potentielles aquifères* déterminées dans la partie aval de la plaine alluviale de la Dore

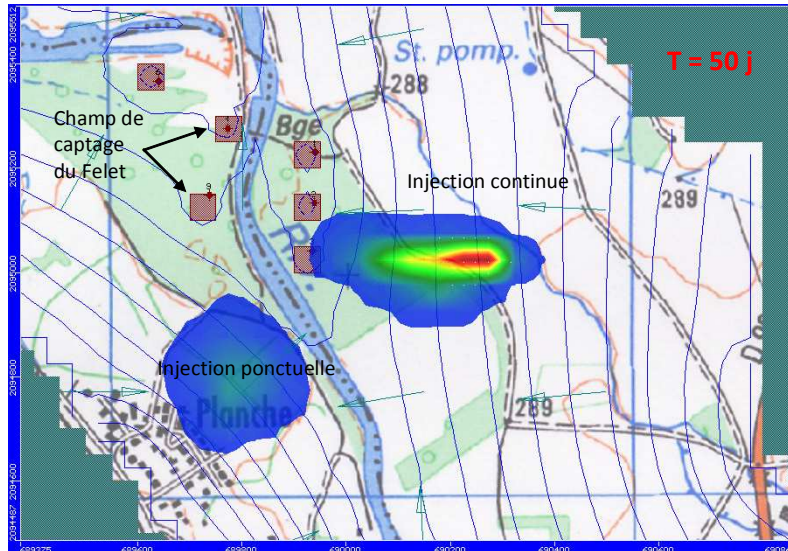


Figure 4. Exemple de simulation de panaches de deux substances chimiques conservatives au niveau du champs de captage AEP du Felet dans un secteur identifié à fort potentiel aquifère (zone 4). Les lignes grises représente les isopièzes (équidistance 0.5m) et les flèches les directions d'écoulement de la nappe libre.

6. Discussion, limites de la méthode et perspectives d'amélioration

L'évaluation des potentialités aquifères peut être complétée par l'optimisation des débits à prélever dans la nappe. Il est possible de faire fonctionner avec Modflow 4.3 (Modular Groundwater Optimizer) un module d'optimisation des prélèvements d'eau fondé sur le principe des algorithmes génétiques pour optimiser d'une part le nombre de puits à forer pour les 5 secteurs identifiés à fort potentiel aquifère, et d'autre part les débits à prélever ainsi que la localisation exacte des puits.

Il est nécessaire pour cela de formaliser plusieurs contraintes. Pour le cas d'étude, la demande globale à satisfaire serait de 14,3 Millions de m³/an pour l'eau potable. Pour chaque puit, le débit minimal sera fixé à 20 m³/h (en dessous de cette valeur les pompes présentent de très mauvais rendements) avec un rabattement maximal admissible de 1/3 la hauteur d'eau mouillée.

Les critères et sous-critères d'évaluation des secteurs à bon potentiel fait pour l'instant l'objet d'un traitement par la somme pondérée. D'autres techniques permettant de s'affranchir des effets de compensation d'un critère conduisant à une mauvaise note par un autre pour lequel la note serait très élevée pourrait être

introduites afin de prendre en compte les concordances ou discordances entre critères (Electre-Tri ou Prométhée) [ROY 85].

Plusieurs inconvénients subsistent :

- le format raster qui oblige à travailler avec une taille de maille fixée ici à 50 m,
- la résolution du modèle hydrodynamique (fixée également à 50 m) qui, même s'il constitue un bon compromis entre le temps de calcul/ préparation des données, n'est sans doute pas suffisante pour la dernière tâche, notamment la détermination des cones de rabattement autour des puits.

7. Conclusion

L'objectif de cette recherche était de proposer une méthodologie visant à évaluer les zones de bon potentiel aquifère des têtes de bassin versant. Cette recherche a permis de valider cette approche intégrée associant SIG, AMC et modélisation hydrodynamique. L'approche peut être appliquée à d'autres sites d'étude en milieu alluvionnaire.

Un premier ensemble de secteurs à fort potentiels a pu être identifié même si des traitements complémentaires sont indispensables pour les préciser. Enfin l'incorporation dans la méthode d'un processus d'optimisation pour la localisation précise des captages et la détermination des débits d'exploitation les plus adaptés devrait faire converger l'ensemble de la méthode vers un véritable outil d'aide à la décision en matière de gestion des ressources en eau.

8. Bibliographie

Castany G., 1986 – Principes et méthodes de l'hydrogéologie, Dunod Université, Paris, 236p.

Conservatoire des Espaces et Paysage d'Auvergne, CRENAM, Juillet 2008 – Etude globale de la plaine alluviale de la Dore. Rapport cartographique pour le Parc Naturel Régional Livradois Forez

Cubizolle H., Octobre 2007 – Fonctionnement hydrologique et géomorphologique de la Dore dans sa traversée de la Limagne Orientale (bassin de l'Allier, Massif Central). CRENAM – UMR 5600 CNRS Université Jean Monnet, St-Etienne, 72 p.

Kinzelbach W., 1986 – Groundwater Modelling. An Introduction with Sample Programs in Basic, Elsevier, Amsterdam, 33p.

Lallemand-Barres A. ET Roux J.-C., 1988 – Guide méthodologique d'établissement des périmètres de protection des captages des eaux souterraines destinées à la consommation humaine, Edition BRGM, Orléans, 287p.

- Mimoun D., Graillot D., 2008 – Etude des potentialités aquifères de la nappe alluviale dans la basse vallée de la Dore Rapport d'étude pour le Parc Naturel Régional Livradois Forez – Ecole des Mines de Saint-Etienne, 66 p
- Roy B., 1985 – Méthodologie multicritère d'aide à la décision, Economica, Paris
- Saaty T., 1984 Décider face à la complexité : une approche analytique multicritère d'aide à la décision. Coll. Université Entreprise, Entreprise Moderne d'Édition.

Cartographie de l'occupation du sol pour la modélisation hydrologique spatialisée du cycle de l'eau en zone périurbaine

David Béal* — **Matthieu Gagnage*** — **Christine Jacqueminet*** —
Saïda Kermadi* — **Kristell Michel*** — **Sonja Jankowsky**** — **Flora Branger**** — **Isabelle Braud****

* *Université de Lyon, UMR CNRS EVS,
18, rue Chevreul, F-69 364 Lyon cedex
Christine.Jacqueminet@univ-st-etienne.fr*

** *Cemagref, UR HHLV
CP220, 3bis Quai Chauveau, F-69336 Lyon cedex 9*

RÉSUMÉ. *Les bassins-versants péri-urbains soumis à un changement rapide de l'occupation du sol, sont exposés à l'accroissement des risques de pollution et d'inondations. La prise en compte de l'hétérogénéité de la surface (coexistence de zones rurales et urbanisées) et des chemins de l'eau naturels et artificiels, dans une simulation en continu sur quelques dizaines d'années de l'hydrologie de ces zones, est l'une des questions-clé du projet AVuPUR (Assessing the Vulnerability of Peri-Urban Rivers). Ceci requiert entre autres, une cartographie de l'occupation du sol qui est menée à différentes échelles à l'aide d'images aériennes et spatiales à très haute résolution. Des méthodes manuelles et semi-automatiques d'extraction sont développées sur ces différents types d'images. Les contraintes liées à l'introduction de ces données dans des modèles hydrologiques spatialisés sont analysées.*

ABSTRACT. *Peri-urban basins, face a quick land-use change, increasing pollution and flood risks. Taking into account the surface heterogeneity (mixture of rural and urbanised areas), and also the natural and artificial water pathways in a continuous and long term modelling of the hydrological cycle in peri-urban areas, is one of the focus of the AVuPUR (Assessing the Vulnerability of Peri-Urban Rivers) project. This requires, among others, a land cover mapping at various scales. The latter is performed by using aerial photography and very high resolution satellite imagery. Manual interpretation and semi automated methods are developed on both images sources and compared. Constraints due to mapping data integration in distributed hydrological models are analyzed.*

MOTS-CLÉS : *modèle hydrologique spatialisé, occupation du sol, télédétection, images très haute résolution, approche orientée objet*

KEYWORDS: *distributed hydrological model, land cover mapping, remote sensing, very high resolution, image segmentation, object-oriented image analysis*

1. Introduction

Le développement de l'urbanisation et les pollutions qui y sont associées augmente la pression sur les eaux des rivières périurbaines ainsi que le risque d'inondations. Par ailleurs, afin de respecter les contraintes posées par la Directive Cadre sur l'Eau, il est nécessaire de disposer d'outils diagnostic et prédictifs pour quantifier l'impact de l'urbanisation sur le cycle hydrologique et à terme sur la qualité des eaux, mais aussi pour proposer des solutions d'aménagement permettant de respecter et/ou d'améliorer le bon état écologique des cours d'eau. Dans le cadre du projet AVuPUR (Assessing the Vulnerability of Peri-Urban Rivers) qui porte sur l'étude de la vulnérabilité des hydro-systèmes périurbains soumis à un accroissement de l'urbanisation et aux changements climatiques, l'objectif est d'améliorer les outils de modélisation hydrologique spatialisée pour mieux prendre en compte les spécificités des milieux périurbains et notamment l'impact des surfaces imperméabilisées, mais aussi des éléments anthropiques qui viennent perturber la dynamique des écoulements (routes, chemins, fossés, réseaux d'assainissement).

Dans le modèle hydrologique, le découpage possède deux niveaux : celui des sous-bassins versants et celui des hydropaysages selon la méthode proposée par Dehotin et Braud (2008), qui consiste à croiser des informations d'occupation du sol, de pédologie, de géologie et de pente. A l'échelle de petits bassins de quelques km^2 , il s'agit de modéliser de manière couplée les fonctionnements des différents éléments hydrologiques représentés de manière explicite (parcelles agricoles, parcelles urbaines, routes, haies, réseaux etc..) et leurs interactions. A l'échelle d'un bassin de 150 km^2 , leur fonctionnement doit être intégré à l'échelle des sous-bassins versants ruraux ou urbains par des modélisations de type réservoir et/ou s'appuyant sur la représentation des pourcentages de flux associés à différentes occupations du sol. Pour mettre en place cette modélisation, notamment sur les petits bassins versants, une cartographie détaillée de l'occupation des sols est nécessaire.

Cette communication porte donc sur la construction d'une information de type occupation du sol, spécifique à la problématique de l'écoulement des eaux, extraite d'images aériennes et spatiales de résolution et de date différentes. L'évolution de l'occupation du sol est étudiée de 1945 à 2008 à l'aide de photographies aériennes N/B et couleurs et d'images satellites à haute et très haute résolution. Le bassin étudié est le bassin de l'Yzeron (150 km^2) s'étendant des Monts du Lyonnais à la périphérie ouest de la ville de Lyon.

Les objectifs et les défis méthodologiques sont les suivants :

- Déterminer la résolution d'image la plus appropriée pour rendre compte des propriétés de perméabilité des surfaces et du cheminement des eaux : photographies aériennes et images satellites à très haute résolution sont comparées ;
- Mettre au point une méthodologie d'extraction de l'occupation du sol, la moins coûteuse possible en temps de travail eu égard à l'importance des données traitées et à la superficie du bassin versant étudié ;
- Fournir des informations sur l'occupation du sol adéquates aux différentes échelles de modélisation : une information très détaillée à l'échelle de sous-bassins versants pilotes et une information plus globale à l'échelle de l'ensemble du bassin versant de l'Yzeron ;
- Fournir une information dans un format compatible avec la modélisation hydrologique.

Ce projet étant en phase de réalisation, nous présenterons l'élaboration d'une typologie d'occupation du sol à des fins de modélisation hydrologique et la constitution du jeu de données images ; aux deux niveaux de modélisation, les différentes méthodes de cartographie manuelle et semi-automatique conduites sur différentes images et enfin, les contraintes d'intégration des cartes réalisées, posées par la modélisation hydrologique spatialisée.

2. Elaboration d'une typologie de l'occupation du sol et choix des données images

Trois grandes catégories de surfaces occupent le BV de l'Yzeron : les espaces forestiers, les espaces agricoles et les espaces urbains et périurbains. Dans chacune de ces catégories, nous différencions les surfaces imperméables, telles que les bâtiments, la voirie routière et les recouvrements de bitume ou de béton, et les surfaces perméables, représentées par les espaces végétalisés et les sols à nu ou couverts d'un recouvrement poreux (graviers...). Afin de renseigner sur les propriétés d'infiltration et d'évapotranspiration des couverts végétaux, nous individualisons au sein du parcellaire agricole, les cultures permanentes (prairies, vergers et cultures en rang) et les cultures annuelles. Dans les espaces forestiers, les peuplements à dominance de conifères et les peuplements à dominance de feuillus sont également distingués. Les discontinuités de l'occupation du sol sont restituées par les structures linéaires du paysage (voirie, haies, limites du parcellaire agricole et foncier) et les limites des objets eux-mêmes (bâti, couvert végétaux...).

Plusieurs contraintes prévalent au choix des données images. La distinction des surfaces perméables et imperméables au sein d'entités d'extension réduite et hétérogènes, telles que les espaces urbanisés, ainsi que la restitution des discontinuités fines du paysage, ont imposé l'utilisation d'images à très haute résolution. La reconnaissance des cultures annuelles et des cultures permanentes nécessite l'utilisation d'images d'une même année enregistrées à des saisons différentes (fin du printemps et fin d'été). L'apport d'une information dans la bande

du proche-infrarouge (données satellites) contribue à la caractérisation de l'activité agricole et permet la discrimination des peuplements forestiers.

Un jeu de données images a été constitué sur la période allant de 1945 à 2008 avec un pas de temps de 10 ou 20 ans (Tableau 1). La BDOrtho IGN a été choisie comme référentiel géographique. En 2008, plusieurs types d'images à très haute résolution ont été acquises afin de pouvoir déterminer la résolution spatiale et spectrale la plus appropriée. Les images Spot acquises en 1990 et 1999 permettront de reconnaître dans le parcellaire agricole, les cultures annuelles et permanentes.

2008	1999	1990	1970	1945
BDOrtho IGN R : 0.5 m	BDOrtho IGN	Photographies aériennes N/B	Photographies aériennes N/B	Photographies aériennes N/B
Quickbird R : 2.4 m				
Spot R : 2.5 m	Spot R : 10 et 20 m	Spot R : 10 et 20 m		

Tableau 1. Les données images du projet AVuPUR : en gras, les images dont les traitements sont développés dans cette communication (R : résolution)

A chacun des deux niveaux de modélisation, l'exploitation de ces images donne lieu au développement de méthodologies de traitements d'image et aboutit à l'élaboration de typologies propres de l'occupation du sol. (Tableau 2).

3. Premier niveau de modélisation hydrologique : cartographie détaillée de l'occupation du sol sur deux sous-bassins versants pilotes

A ce premier niveau de modélisation hydrologique, la parcelle cadastrale constitue l'unité de base sur laquelle les processus hydrologiques sont simulés. La modélisation intègre également les éléments perturbateurs de la dynamique de l'eau, comme les fossés ou les haies. A cette échelle du sous-bassin versant, l'information sur l'occupation du sol doit être la plus exhaustive et détaillée possible eu égard aux potentialités des images. Ainsi, si les haies sont cartographiées, la restitution exhaustive des fossés sur photographies aériennes s'est révélée impossible. L'information sur les fossés a été collectée par des relevés de terrain.

Echelle de modélisation hydrologique	Type et résolution d'images	Méthodologie traitements d'image	Typologie de l'occupation du sol	Mode d'intégration dans la modélisation
Ensemble du bassin versant	Quickbird 2.5 m	Traitement Semi-automatique	Globale et simplifiée (8 thèmes)	Pourcentage de surfaces par ss-BV
	BDortho 0.5 m	Traitement Semi-automatique	Globale et simplifiée (6 thèmes)	Pourcentage de surfaces par ss-BV
Sous-bassin versant	BDortho 0.5 m	Digitalisation manuelle	Détaillée (20 thèmes)	Objets élémentaires surfaciques
		Validation des traitements numériques		

Tableau 2. Images et traitements réalisés aux deux niveaux de modélisation

Cette cartographie réalisée à partir de la BDOrtho IGN, s'est appuyée sur le cadastre pour définir les limites des parcelles foncières. La reconnaissance des objets sur photographies aériennes en couleurs naturelles à 0.5 m de résolution est aisée. Par contre, on constate une réelle difficulté à distinguer en milieu urbain, les matériaux de recouvrement poreux (graviers) des matériaux imperméables (bitume, toitures). A cette résolution, la reconnaissance de ces objets repose sur leur couleur qui peut être similaire. Leur texture n'est pas perceptible.

Les sous-bassins versants cartographiés à cette échelle représentent une superficie totale d'environ 11 km². Leur cartographie selon une typologie détaillée a nécessité 2 à 3 mois de digitalisation à temps plein. Cette méthode est très coûteuse en temps et ne peut être envisagée pour la totalité du bassin versant de l'Yzeron à différentes dates. Cette cartographie détaillée est utilisée pour la validation des traitements automatiques à réaliser sur l'ensemble du bassin versant.

4. Deuxième niveau de la modélisation hydrologique

A l'échelle de l'ensemble du bassin versant, la modélisation hydrologique se fera par sous-bassin, en tenant compte des pourcentages occupés par les différentes occupations du sol, notamment les zones perméables et imperméables et/ou les zones connectées ou non au réseau hydrographique naturel. La grande superficie du bassin versant de l'Yzeron impose le développement de méthodes automatisées d'extraction de l'occupation du sol. Deux chaînes de traitement ont été envisagées en se basant dans un premier temps sur une segmentation multi-canaux/multi-résolution et ensuite sur une classification supervisée à l'aide de paramètres radiométriques et géométriques. Elles se différencient d'une part à l'étape de la segmentation et d'autre part par le nombre de paramètres utilisés. La segmentation

des ortho-photographies résulte de la fusion des segmentations de divers masques. Celle de l'image Quickbird utilise la bande du proche-infrarouge en plus de celle du visible.

4.1. Cartographie globale par traitement semi-automatique de la BDOrtho IGN

Dans les limites du bassin versant de l'Yzeron, la BDOrtho IGN se compose de 217 dalles de 1km². La résolution spatiale est de 0.5 m et les images sont en composition colorée RVB dans le spectre du visible. L'objectif est d'automatiser le plus possible la méthode. Les classes d'occupation du sol sont les suivantes : végétation arborée (apparaissant vert sombre à l'image), végétation herbacée ou culture (vert clair), chemins non bitumés ou sol nus (blanc/rouge), routes bitumées (gris), toitures (rouge, blanc, gris voire vert), piscines (bleu) et étendues d'eau (vert).

L'algorithme, développé à l'aide du logiciel Matlab, réalise une classification orientée objet en utilisant la dimension pseudo-spectrale (combinaisons des niveaux de rouge, vert et bleu) et la dimension spatiale (indices de texture) de l'image. Classiquement, la première étape est une segmentation de l'image qui consiste à découper l'espace en objets (polygones). La seconde étape correspond à une classification, c'est-à-dire attribuer à chaque objet une classe en faisant des statistiques sur les valeurs des pixels du polygone. Les étapes de l'algorithme sont illustrées Figure 1. Cinq masques sont successivement extraits :

- le masque Ombre, issu d'un seuillage sur les trois couleurs, individualise les ombres ;
- le masque Vert, la végétation et les étendues d'eau ;
- le masque Rouge & Gris, les objets minéraux tels que les routes, les toits, les serres, les sols nus. La distinction en deux classes n'est pas faite car les objets minéraux sont constitués à diverses proportions de ces deux classes ;
- le masque Bleu, les piscines ;
- un dernier masque est créé à partir d'un indice de texture. Il permet de distinguer les objets homogènes comme les sols nus, les parcelles couvertes de végétation herbacée ou les étendues d'eau. *A contrario*, les objets non homogènes sont les forêts (mélange de pixels verts plus ou moins sombres à cause du volume de la canopée) ou les routes (bruit provenant des bandes blanches, voitures, poussières de bordures) ou toits (bruit provenant de l'inclinaison et des orientations différentes des pans).

Chacun des masques subit un filtrage (on ne retient que les pixels ayant huit voisins de la classe) et une détection de contours (par calcul de l'indice de texture 3x3). Les contours sont fusionnés et un polygone est attribué à chaque objet.

Puis, des statistiques simples sont faites dans chaque polygone ce qui permet de déterminer la classe la plus représentée.

Des confusions classiques persistent cependant (20%) entre les types de végétation et entre les sols nus et les routes ou les toits. De plus, les étendues d'eau, de même couleur que la végétation, doivent être systématiquement classifiées manuellement.

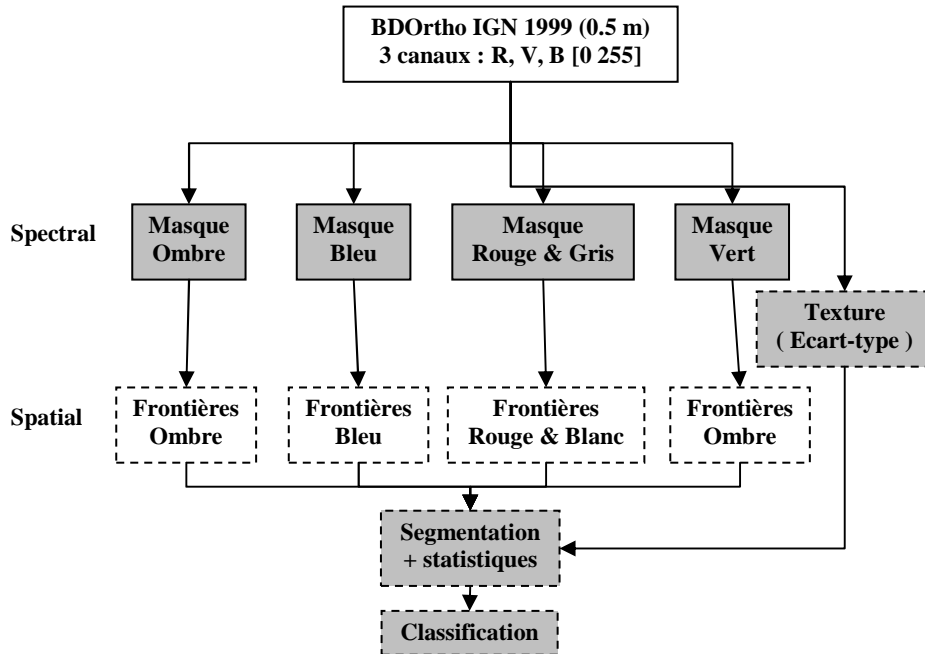


Figure 1. Schéma de l'algorithme semi-automatique de cartographie de l'occupation du sol à partir de la BDOrtho IGN

4.2. Cartographie globale par traitement semi-automatique d'une image satellite Quickbird

L'image satellite Quickbird du 29/08/2008, enregistrée dans le spectre du visible et du proche-infrarouge, a une résolution de 2.4 m. Elle est classifiée selon l'approche orientée objet à l'aide du logiciel Definiens Developer. Le processus de cartographie se base à la fois sur la segmentation multirésolutions de l'image et sur la classification des objets issus de cette segmentation. La segmentation est étalonnée à partir de paramètres de couleur, de forme et de taille, pour chaque objet à extraire. Le logiciel découpe l'image en polygones, délimitant ainsi des objets reconnaissables à l'œil humain (parcelles, routes...).

L'étape de classification hiérarchique de l'ensemble des polygones inclut ensuite divers paramètres discriminant pour chaque type de surface au sol. Ainsi, des paramètres de forme, de taille des polygones, et des variables issues d'outils plus classiques de la télédétection (indices de végétation, de texture, valeurs statistiques des canaux) ont été utilisés. Aux différents niveaux de la procédure hiérarchique, sont extraits les classes d'occupation du sol (Karsenty *et al*, 2007) (Figure 2).

Une fois le processus automatique achevé, un certain nombre d'erreurs ont été identifiées (confusion ombre et végétation sombre, routes et bâti), et subsiste le

problème des polygones d'ombre. Une vérification systématique a donc été menée sur l'ensemble de l'image pour corriger ces erreurs et attribuer manuellement aux polygones de la classe ombre, un type d'occupation du sol.

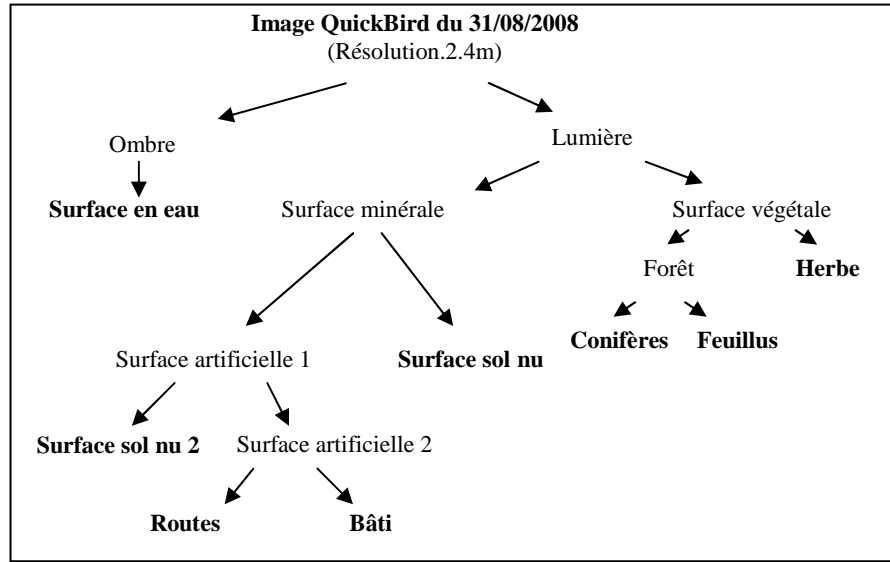


Figure 2. Classes d'occupation du sol extraites (en gras) aux différents niveaux de segmentation de l'approche orientée objet

6. Les contraintes d'intégration des cartes d'occupation du sol dans les modèles hydrologiques spatialisés

La modélisation à l'échelle des sous-bassins versants est effectuée avec la plateforme de modélisation orientée objet LIQUID (Viallet *et al.*, 2006). Un modèle dans la plateforme se compose de plusieurs modules décrivant des processus hydrologiques différents, s'appliquant sur les hydro-paysages. Ceux-ci sont pris en compte sous forme de polygones ou de ligne pour la rivière. Certains éléments linéaires, comme les haies et routes, sont également considérés sous forme de polygones afin de réaliser des bilans hydrologiques impliquant leur surface. Le choix des modules dépend principalement de l'occupation du sol. Ceci est lié à la nature des modules existant dans la plateforme, qui décrivent entre autres explicitement les haies, l'écoulement en surface, l'infiltration dans le sol, la rivière et les parcelles urbaines. Les zones urbaines sont modélisées par le module URBS (Rodriguez *et al.*, 2008), qui s'appuie sur une parcelle cadastrale sous forme vectorielle, ayant l'étendue des surfaces perméables et imperméables (bâtiment, parking et moitié de rue) comme attribut. Afin de pouvoir utiliser la carte

d'occupation du sol comme critère principal du maillage, quelques contraintes doivent être respectées. Premièrement, le modèle exige une couche vectorielle avec une topologie correcte impliquant le respect des règles suivantes: aucun chevauchement des polygones, pas d'espace vide entre les polygones, correspondance des sommets des polygones adjoints et pas de multi-polygones. Il ne faut pas non plus avoir des polygones « troués » car il n'est pas possible d'avoir un module encastré dans un autre. Dans ce cas, un décalage des surfaces intérieures (trous) vers les bords du polygone englobant par échange des surfaces ou un redécoupage des polygones « troués » devient nécessaire (voir Figure 3). De plus, il faut prendre en compte le nombre et la taille des polygones. Comme chaque polygone est le support d'un module connecté aux autres polygones par des interfaces, qui constituent elles-même un module, le modèle devient vite assez complexe s'il y a un grand nombre de polygones ou des polygones trop petits. Par conséquent, une simplification de la carte d'occupation du sol, ainsi que des autres couches prises en compte, est nécessaire, tout en conservant les caractéristiques représentant les processus hydrologiques. Pour les deux sous-bassins versants, d'une taille d'environ 5 km² chacun, on a choisi de ne considérer que les polygones d'une surface plus grande que 10 m². Un dernier critère est de conserver la connectivité des polygones, par exemple d'éviter qu'un polygone de route soit coupé par le feuillage des arbres au bord de la route (Figure 3).



Figure 3. Extrait de la carte détaillée de l'occupation du sol : à gauche, l'image initiale, au milieu et à droite, les images de polygones corrigés. La première correction opère le découpage d'un bois (en trame façonnée) avec la route (en blanc) pour assurer la connectivité de la route. Pour supprimer les trous dans la zone arborée, la première méthode (image du milieu) consiste à fusionner les trous et à découper la zone arborée. Cette méthode est assez coûteuse en temps de travail. Une deuxième possibilité, moins coûteuse en temps, est montrée à droite : on supprime les trous. Le découpage de la zone arborée avec la route crée des polygones en plus. Ceux-ci peuvent être effacés (image de droite), ou être fusionnés pour réduire leur nombre, tout en gardant à peu près la surface arborée (image du milieu).

7. Conclusions et perspectives

Le travail d'ores et déjà réalisé sur la cartographie de l'occupation des sols à différentes échelles montre la complémentarité des différentes sources d'informations. La haute résolution des images permet de distinguer au sein des parcelles bâties, les surfaces perméables et imperméables bien qu'il subsiste toujours un biais dans la discrimination des caractéristiques de perméabilité des surfaces minérales. Ces confusions persistantes peuvent être dans un premier temps réduites en ajoutant des paramètres radiométriques et/ou géométriques. L'apport de l'information dans le proche infrarouge de l'imagerie satellitale améliore la caractérisation des couverts végétaux, une caractérisation complétée par l'utilisation d'images enregistrées à des saisons différentes. Ce travail montre qu'il est possible d'orienter la méthodologie du traitement d'images, manuelle ou semi-automatique, à partir des besoins de la modélisation hydrologique à différentes échelles. L'intégration de cette information au format vectoriel, dans les modèles hydrologiques nécessite des post-traitements manuels (polygones trop petits ou emboîtés, objet discontinu). Ces post-traitements peuvent vraisemblablement être automatisés en utilisant les propriétés géométriques des polygones fournies par les algorithmes utilisés. Une fois le modèle mis au point sur les données récentes, les cartes d'occupation antérieures reconstituées permettront d'inférer l'impact de sa modification sur le cycle hydrologique.

Remerciements

Le projet AVuPUR est financé par l'ANR 2007 «Vulnérabilité Milieux Climat ». Les images Spot ont été acquises dans le cadre d'un contrat ISIS.

Bibliographie

- Dehotin, J., Braud, I., «Which spatial discretization for distributed hydrological models? Proposition of a methodology and illustration for medium to large scale catchments». *Hydrology and Earth System Sciences*, vol. 12, 2008 p. 769-796.
- Karsenty A, Antunes A.F.B., Centeno J.S., «Classification orientée objet de la perméabilité des sols en zone urbaine à l'aide d'imagerie très haute résolution et de données laser scanner à Curitiba (Brésil)», *Anais XIII Simpósio Brasileiro de Sensoriamento Remoto*, Florianópolis, Brasil, 21-26 avril 2007, INPE, p. 565-572.
- Rodriguez F., Andrieu H., Morena F., «A distributed hydrological model for urbanized areas – Model development and application to case studies », *Journal of Hydrology*, vol. 351, 2008, p. 268-287.
- Viallet P., Debionne S., Braud I., Dehotin J., Haverkamp R, Sâadi Z., Anquetin S., Branger F., Varado N. (2006). « Towards multi-scale integrated hydrological models using the LIQUID framework », 7th International Conference on Hydroinformatics 2006, Nice, France, 4-8 September 2006, vol. I, p. 542-549.

Modélisation pour la valorisation des continuums écologiques départementaux dans l'aménagement du territoire

Application aux Scot de la Loire (42)

Mireille Batton-Hubert* — Marie Bonneville** — Thierry
Joliveau*** — Frédéric Paran*

* Centre SITE, Ecole Nationale Supérieure des Mines de Saint-Etienne
158, cours Fauriel, 42023 Saint-Etienne Cedex 2

batton@emse.fr

paran@emse.fr

** Conservatoire des Espaces Naturels du Languedoc-Roussillon
Montpellier

mariebonneville@hotmail.com

*** CRENAM, Université Jean Monnet de Saint-Etienne
Saint-Etienne

thierry.joliveau@univ-st-etienne.fr

RÉSUMÉ. L'article présente une expérience de recherche conduite depuis 2007 sur le thème général des Infrastructures Vertes et Bleues (IVB). Cette démarche a été initiée par un organisme d'Etat pour conduire les différentes parties intéressées à mieux prendre en compte les enjeux écologiques dans les Schémas de Cohérence Territoriale (SCoT) et documents de planification territoriale à l'échelle des agglomérations.

ABSTRACT. This paper presents a research experiment started in 2007 about the French "blue and green infrastructures". At the prompting of French local administration, the aim of this experiment consists in taking into account involved actors and ecological stakes in local territorial plan at the conurbation scale.

MOTS-CLÉS: continuums écologiques, planification territoriale, analyse multicritère, Sig

KEYWORDS: ecological continuum, territorial plan, multicriteria analysis, Gis

1. Introduction

Les démarches formelles habituelles de la planification territoriale se déroulent généralement en deux grandes étapes : la première s'oriente vers la compilation de documents préparatoires des servitudes du territoire, et la seconde vers une discussion/négociation en de multiples commissions de la conformité des actions de développement avec les objectifs environnementaux. La démarche IVB (Infrastructures Vertes et Bleues) a l'ambition de donner des fondements scientifiques à la planification territoriale. Elle est fondée sur la prise en compte des apports théoriques de l'écologie du paysage et de données traitées à l'aide d'outils géonumériques. Ces outils permettent une représentation cartographique des critères à prendre en compte dans la décision. Ce travail fait suite à la demande d'un service de l'Etat, la Direction Départementale de l'Équipement, (Dde-Loire) soucieux de rendre sa démarche plus efficace et plus participative dans le cadre de l'élaboration des Scot (Schéma de Cohérence Territoriale).

Cette demande est structurée par 4 enjeux majeurs : 1) intégrer une dimension écologique, 2) établir une procédure de planification territoriale, 3) assurer le traitement géonumérique des données et la combinaison multicritère, 4) intégrer une dimension participative. Il va s'agir de délaissier le champ de l'expertise pour proposer un dispositif spécifique à l'interface de la recherche et de l'action pour construire la question posée. Ce travail a abouti à la mise en place d'un atelier de formation/recherche mettant en œuvre la méthode sous forme d'un jeu de rôles. Il a été testé avec des chercheurs et des praticiens dans le cadre des ateliers Sageo⁸ (Batton-Hubert *et al.*, 2007a), dont le bilan a été présenté lors des conférences Opde⁹ (Batton-Hubert *et al.*, 2008), et d'un cours de Master-Recherche (février 2008 et 2009). La démarche a aussi été valorisée lors d'une communication locale en partenariat avec la Dde-Loire (Batton-Hubert *et al.*, 2007b).

Notre exposé suivra le plan suivant. D'abord, nous évoquerons le contexte de la démarche IVB dans les Scot. Ensuite, nous rappellerons de manière synthétique la démarche méthodologique de modélisation des IVB sur le Sud-Loire. Enfin, nous présenterons les éléments d'enrichissement de la méthode sur le Nord-Loire, ainsi que les principaux questionnements et perspectives.

2. La démarche IVB dans les SCoT

En amont de la démarche IVB, il existe des précurseurs. On peut citer les travaux réalisés en Suisse (Berthoud et Righetti, 2004), ou en France pour le Réseau

⁸ Colloque International de Géomatique et d'Analyse Spatiale, atelier « Modélisation spatiale et décision territoriale participative », juin 2007, Saint-Etienne, Clermont-Ferrand, France.

⁹ Les Outils Pour Décider Ensemble, juin 2008, Université de Laval, Québec.

Ecologique Départemental de l'Isère (Berthoud, 2001) et pour les continuums biologiques de l'agglomération de Belfort-Montbéliard (Asconit, 2003).

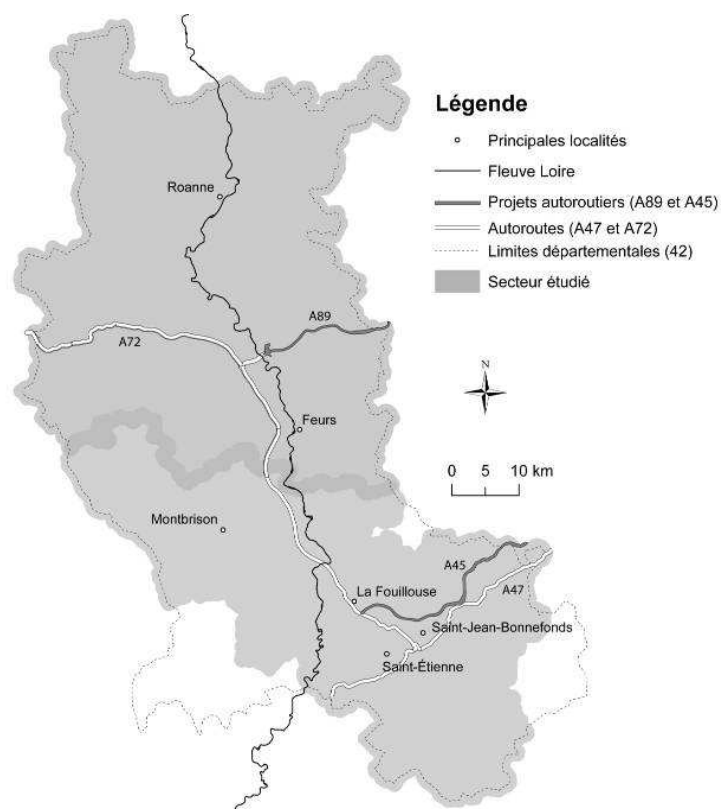


Figure 1. Localisation de la zone d'étude : Scot Sud-Loire (Saint-Etienne, Montbrison) et secteur Nord-Loire (Roanne, Feurs)

2.1. Des enjeux nationaux aux enjeux locaux

Le concept d'IVB a été proposé au niveau régional par la Diren (Direction Régionale de l'Environnement) Rhône-Alpes (Asconit, 2005) pour intégrer les objectifs de préservation de la biodiversité, de la qualité paysagère, des aménités, dans les démarches de planification territoriale. Au niveau national, le « Grenelle de l'environnement » a abordé la question en termes de « trames vertes et bleues » envisagées comme un instrument décentralisé d'aménagement durable, de concertation et de gestion intégrée du territoire. Au niveau local, celui de l'agglomération, il y a l'expérimentation menée sur le Scot Sud-Loire (Asconit,

2005a et 2005b) à la demande de la Dde, la Diren et la Ddaf (Direction Départementale de l'Agriculture et des Forêts).

L'application présentée ici est une expérimentation ultérieure réalisée à la demande de la Dde-Loire. Elle est dédiée pour une part au Scot Sud-Loire et pour une autre part à la préparation des Scot dans le Nord-Loire (Figure 1.). L'objectif principal consiste à aider, à améliorer la méthodologie de construction des IVB dans les exercices de planification. En effet, les méthodes permettant d'objectiver les IVB sont loin d'être totalement stabilisées. Il va s'agir d'assurer le passage d'une expérimentation décentralisée descendante et technocratique à une expérimentation ascendante et participative s'appuyant sur la valorisation du potentiel local d'innovation et de recherche.

2.2. Définitions : des IVB dans les Scot

Les Infrastructures Vertes et Bleues¹⁰ sont définies comme un ensemble d'espaces reliés et hiérarchisés intégrant les déplacements doux des hommes et les grands axes de déplacement des animaux ou continuums écologiques. L'enjeu est d'intégrer de tels réseaux lors de l'élaboration des Schémas de Cohérence Territoriale visant à la coordination dans l'espace et dans le temps des différentes politiques publiques d'aménagement du territoire à l'échelle intercommunale. Il s'agit d'un projet de territoire qui détermine les grands équilibres entre les espaces urbains, à urbaniser et les espaces naturels, agricoles et forestiers, dans une perspective de développement territorial durable.

3. Modélisation des IVB

Cette partie résume les différents aspects de la méthode initiale (Asconit, 2005a et 2005b) et de la méthode de modélisation issue des expérimentations sur le Sud-Loire (Bonnevialle, 2007 ; Batton-Hubert *et al.*, 2008). Au final, le territoire étudié couvre la quasi-totalité du département de la Loire (Figure 1.).

3.1. Méthode initiale

La construction des continuums écologiques est fondée sur la simplification de la notion de réseau écologique (Berthoud, 2001). Sur la zone d'étude, on distingue 4 types de continuums, dans lesquels se déplacent les animaux, associés à des espèces emblématiques : Boisé de Basse Altitude (Bba, chevreuil), Milieux Aquatiques et Humides (Mah, amphibiens), Zones Thermophiles Sèches (Zts, orthoptères), Agricoles Extensifs et des Lisières (Ael, perdrix). La construction de chacun des

¹⁰ « Vert » pour les milieux naturels terrestres et « Bleu » pour les milieux naturels aquatiques.

continuums est effectuée par analyse spatiale sous Sig en mode vecteur par agrégation des données d'occupation du sol Corine Land Cover (2000) dans les 4 groupes de milieux précédemment cités. La spatialisation d'un continuum est obtenue selon la formule [1].

$$c = rd \quad [1]$$

Connaissant la résistance (r) d'un milieu, le potentiel initial de déplacement des animaux fixé à 3000 points (c), il est possible de calculer la distance (d) qu'une espèce peut théoriquement parcourir sur un territoire et de la projeter sur une carte pour obtenir le continuum potentiel. La méthode initiale utilise 4 classes de coefficients de résistance (r), identiques pour les déplacements de tous les animaux considérés, et définies selon la structuration théorique du réseau écologique (Berthoud, 2001). Les 4 classes de coefficients de résistance sont les suivantes : r=0 pour les milieux structurants (zones réservoirs et de développement des espèces emblématiques), r=5 pour les milieux attractifs (milieux connexes et complémentaires aux milieux structurants où les déplacements sont faciles), r=30 pour les milieux peu fréquentés (milieux connexes aux milieux structurants et attractifs où les déplacements sont plus rares et la pénétration moins facile), r=100 pour les milieux répulsifs (milieux a priori non fréquentés par les espèces emblématiques et représentant un obstacle au déplacement). D'après la formule [1], par exemple, les animaux ne pourront parcourir que 30m dans un milieu qualifié de répulsif.

La carte finale est obtenue par superposition de la carte des continuums potentiels et d'autres critères hiérarchisés¹¹ : milieux remarquables (ex : zones Natura 2000), loisirs, liaisons douces, paysages (ex : chemins de randonnées) et obstacles (ex : routes, zones urbaines).

3.2. Méthode de modélisation et de combinaison de critères pour le Sud-Loire

Le calcul des continuums est désormais effectué en mode raster et utilise une fonction de propagation diffusion de type « cost-distance » (ArcGis 9.2, spatial analyst). Ce mode est plus adapté à la représentation de phénomènes continus, comme la spatialisation des continuums écologiques (déplacement des animaux). Il permet de plus de jouer sur le nombre de points de déplacement limite attribués aux animaux et de tenir compte de toute la gamme dans les représentations (Figure 4.). Par ailleurs, le temps de calcul est considérablement raccourci. Pour les autres thèmes, des modifications ont aussi été apportées, même si les classes de hiérarchisation restent les mêmes. Le critère milieux remarquables a été modifié en

¹¹ Les milieux remarquables et loisirs, liaisons douces, paysages sont hiérarchisées en 3 classes selon leur participation significative (ps), forte (pf), majeur (pm) respectivement en fonction de la sévérité de la réglementation et du degré de protection ou de fréquentation ; les obstacles sont hiérarchisés en 4 classes : obstacle imperméable, gêne incontournable, gêne importante, gêne moindre.

milieux protégés à partir des données de la Diren Rhône-Alpes concernant les zonages de protection réglementaire et contractuelle, et d'inventaire. Le critère loisirs, liaisons douces, paysages a été scindé en deux. Un critère paysage-patrimoine qui est fondé sur l'attractivité, la valeur du site, les zones d'intérêt patrimonial ou paysager, la densité d'habitat et de fréquentation à partir des données de l'Insee¹². Un critère déplacement doux qui reprend le réseau de chemin. Enfin, le critère obstacle n'a pas été modifié.

Cette fois, la carte finale est obtenue par agrégation des critères et couplage Sig-Amc, méthode qui connaît un succès certain dans de multiples problématiques territoriales (Molines, 2003). Parmi, les nombreuses méthodes multicritère la méthode Ahp¹³ développée par (Saaty, 1984) a été retenue. Seuls les critères déplacements doux et obstacles sont superposés. Chaque critère peut-être pondéré à dire d'experts, ou bien à dire d'acteurs impliqués. Une telle approche participative est simulée dans des ateliers en utilisant l'outil « weighted sum » (ArcGis 9.2, spatial analyst). La simulation de la concertation est fondée sur des fiches acteurs établies en s'inspirant de la structure de celles utilisées dans des travaux antérieurs portant sur la simulation d'une négociation (Paran, 2005).

4. Enrichissement de la méthode, questionnement et perspectives

Le bilan tiré de l'expérimentation sur l'ensemble du Scot Sud-Loire et sur des secteurs plus restreints (La Fouillouse, Montbrison, Saint-Jean-Bonnefonds) a conduit à valider l'approche raster et à développer un certain nombre d'améliorations pour l'étude du Nord-Loire (Figure 1.). L'adaptation de la méthode est fondée sur une réflexion concernant les données d'entrée, le mode de calcul (paramétrage) et des expertises de terrain.

Nous avons observé la sensibilité de la méthode aux données initiales (Figure 2.). L'observation montre, sur de petits secteurs tests, que les continuums générés ne sont pas identiques à partir des données Corine Land Cover (Clc, 1/100000^e) et Spot Théma (St, 1/25000^e). Il apparaît qu'une discussion de la classification Clc est nécessaire. L'absence d'une cartographie de l'occupation du sol à vocation écologique sur l'ensemble du territoire est donc problématique pour ce travail.

Le critère obstacles est désormais intégré dans le calcul des continuums par attribution d'un coefficient de résistance aux classes d'obstacles prédéfinies dans la méthode initiale (Figure 3.). Pour le réseau routier, celui-ci est fonction de la densité du trafic et de la largeur. Une autoroute, par exemple, est considérée comme un obstacle imperméable ($r=100$), alors qu'une route avec moins de 2000 véhicules jour est un obstacle de moindre gêne ($r=10$). De plus, des coefficients de résistance

¹² Institut National de la Statistique et des Études Économiques.

¹³ Analytic Hierarchy Process.

hiérarchisés ont été attribués aux voies ferrées, aux lignes électriques et aux barrages selon leur importance, et au tissu urbain selon son caractère continu à discontinu. A ce stade, il est aussi possible d'intégrer les fuseaux de projets autoroutiers A47 et A89 du département (Figure 1.). Dans le calcul en mode raster chaque obstacle est considéré avec une zone tampon dont la taille est proportionnelle à son importance. Pour une autoroute, par exemple, le tampon affecté est de 100m. Pour une petite route de moindre gêne, il est de 20m.

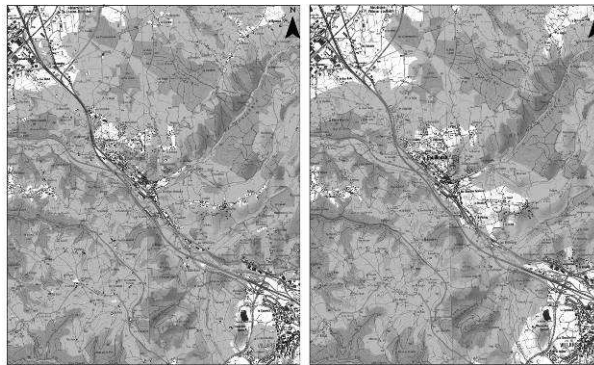


Figure 2. *Potentialités du continuum Bba sur le secteur de La Fouillouse : calculées à gauche avec les données Spot Théma, à droite avec les données Corinne Land Cover (fond : scan 25 Ign)*

L'approche terrain sur de petits secteurs s'est avérée très enrichissante. Elle révèle des informations inexistantes tant sur les couvertures Corine Land Cover que Spot Théma. Des obstacles a priori imperméables peuvent présenter des zones de passage sécurisées (buses et tunnels à faune sous les autoroutes) ou au contraire très dangereuses (absence de grillage, collision). Sur les cours d'eau, nombre de petits obstacles ne sont pas cartographiés (seuils, berges murées, pollution...).

L'intégration de ces obstacles reste toutefois problématique. En effet, une telle notion est relative aux espèces considérées ce qui fait intervenir des problèmes de changement d'échelles et de disponibilité des données. Par exemple, si les murets constituent des obstacles infranchissables pour certains amphibiens, ce n'est pas vrai pour les chevreuils. Dans ce cas, il apparaît que non seulement la cartographie de ce type d'obstacle (murets) n'existe généralement pas et nécessite donc un lourd inventaire de terrain, de plus son intégration à l'échelle d'un Scot reste problématique du fait leur petite taille. L'intégration des obstacles nécessite donc un approfondissement conceptuel et méthodologique.

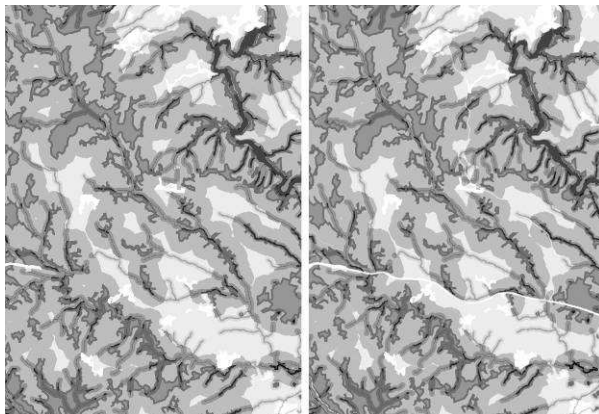


Figure 3. Potentialités pour les continuums écologiques (fortes à nulles du sombre au clair) : à gauche agrégation sans les obstacles, à droite avec les obstacles (données : Corine Land Cover)

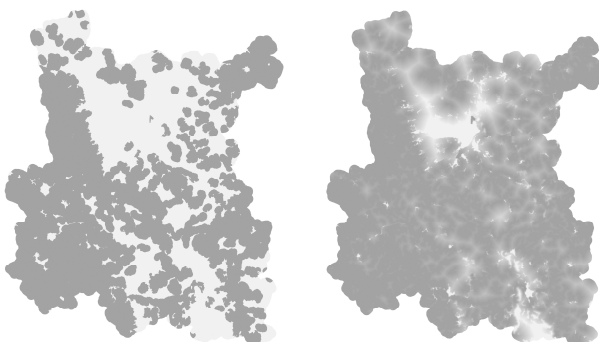


Figure 4. Potentialités du continuum Bba sur le Nord-Loire (fortes à nulles du sombre au clair) : à droite une représentation en 2 classes (0-3000 en sombre, >3000 en clair), à gauche en continue (0 à 195253 points du clair au sombre)

De plus, il conviendrait de mobiliser des recherches plus fondamentales en écologie pour améliorer la modélisation des réseaux écologiques (ex: distinction des espèces, fonctionnement plus réaliste, classification des coefficients de résistance selon les espèces considérées, meilleure prise en compte des obstacles...), de rechercher d'autres méthodes de construction des continuums et d'améliorer les connaissances des populations animales (inventaires, génétique des populations, observations...) pour discuter de l'attribution de 3000 points de déplacement pour l'ensemble des espèces (Figure 4.).

Ce type de travail permettrait un diagnostic fonctionnel des corridors écologiques afin de préciser des périmètres prioritaires et de hiérarchiser les actions et aménagements envisagés. Il peut s'agir de restauration de cours d'eau, de création de points de passage ou de sécurisation de secteurs dangereux, d'amélioration des pratiques agricoles, d'aménagements améliorant la perméabilité du tissu urbain, ou encore d'actions de communication, d'animation ou de sensibilisation. Un tel diagnostic semble d'ores et déjà envisageable à l'échelle d'un Plu.

La démarche IVB peut aussi constituer une approche complémentaire des méthodes visant à cerner les évolutions géohistoriques et paysagères d'un territoire. Un tel travail est engagé par le Centre Site de l'Emse¹⁴ sur l'évolution du régime hydrologique du bassin versant du Lignon du Forez dans le cadre d'un projet lié au Plan Loire Grandeur Nature 3. L'étude sera complétée par une approche statistique des relations pluies-débits et sociologique de la perception des changements à dire d'acteurs.

5. Conclusion

D'une manière générale, les enseignements tirés de cette expérimentation sont les suivants. La méthode IVB constitue une approche complexe et difficilement communicable dont la validité écologique est critiquable faute de données bien adaptées et de validation par des expertises naturalistes sur le terrain. La méthode multicritère proposée nécessite un accompagnement technique, tend à rendre la négociation aspatiale et l'anticipation des conflits difficile. L'atelier quant à lui, s'il est mobilisateur reste très dense avec nombre d'informations à assimiler.

Toutefois, du point de vue du demandeur (Dde-Loire), il s'agit d'une amélioration concrète de la méthode avec un test d'applicabilité direct. Du point de vue des enseignants-chercheurs la méthode et ses limites semblent facilement appropriables. Elle offre la possibilité de tester des hypothèses méthodologiques sur la participation, l'utilisation de l'Amc et la modélisation spatiale. Du point de vue des participants à l'atelier, la démarche reste mobilisatrice, mais abstraite. Elle facilite la transposition dans d'autres lieux et fait apparaître les problèmes potentiels de la méthode en vraie grandeur.

Les perspectives de la démarche sont nombreuses sur le plan des problématiques de recherches et opérationnelles. Dans le domaine des sciences, il est nécessaire d'approfondir ou développer les critères pour la construction d'un modèle de combinaison global, de proposer un modèle pertinent pour les critères patrimoine-paysage et déplacements doux, de revoir le calcul des continuums, de mieux intégrer les obstacles, de mettre au point une stratégie participative d'enrichissement des données et de réfléchir à d'autres approches multicritère. Sur le plan opérationnel, il conviendra d'intégrer des scénarios de développement territorial (urbanisation,

¹⁴ Ecole des Mines de Saint-Etienne.

aménagement...), la notion de changement d'échelle (du Scot au Plu) et de cartographie des zones de conflits potentiels à travers une pondération par secteur.

6. Bibliographie

- Asconit, Etat initial du site et évaluation environnementale du Scot du pays de Montbéliard. Direction Générale de l'Environnement de Franche-Comté, Rapport de synthèse, 2003.
- Asconit, Utilisation d'un système d'information géographique pour l'expression des enjeux de l'Etat, application au territoire du Scot Sud-Loire. Diren ra, Guide technique, 2005a.
- Asconit, Utilisation d'un Sig pour l'expression des enjeux de l'Etat dans le cadre d'un Scot, application au territoire du Scot Sud-Loire. Diren ra, Guide méthodologique, 2005b.
- Batton-Hubert M., Bonnevialle M., Joliveau T., Paran F., « Intégrer une dimension écologique et paysagère dans la planification territoriale - Méthode et questionnement à propos de la démarche Infrastructures Vertes et Bleues (IVB) dans les SCoT de la Loire (France) ». *Actes de la conférence Les outils pour décider ensemble*, Université de Laval, Québec, Juin 2008, 17p.
- Batton-Hubert M., Bonnevialle M., Joliveau T., Paran F., « IVB dans un Scot », *Ateliers, Rencontres internationales Géomatique et Territoire*, Sageo, juin 2007a, 4p.
- Batton-Hubert M., Joliveau T., Bonnevialle M., Paran F., Les méthodes d'identification des Infrastructures Vertes et Bleues (IVB) au service des projets d'aménagement et de développement durable des SCoT. Communication orale, Semaine régionale du développement durable, Préfecture de la Loire, Dde Loire, Ujm-Crenam, Emse-Site, Saint-Etienne, Novembre 2007, 2007b.
- Berthoud G., Righetti A., *Le Réseau écologique National Suisse*, Cahiers de l'environnement n°373 Nature et Paysage, Office Fédéral de l'Environnement des Forêts et du Paysage, 2004.
- Berthoud G., Les Corridors biologiques en Isère : Projet de Réseau écologique Départemental de l'Isère (Redi), Conseil Général de l'Isère, 2001.
- Bonnevialle M., Elaboration d'un outil participatif associant Sig et Amc pour la construction des IVB et favoriser leur intégration dans l'élaboration des Scot - Application au Scot Sud-Loire. Ensm-Se, Université Lille 1, Rapport de master, 2007.
- Molines N., Méthode et outils pour la planification des grandes infrastructures linéaires et leur évaluation environnementale, Mémoire de thèse de doctorat Ujm-Se et Université de Laval au Québec, 2003.
- Paran F., Représentations territoriales pour la gestion équilibrée d'un patrimoine écologique et anthropique dans le domaine de l'eau. Mémoire de thèse de doctorat, Ensm-Se, Ujm-Se, 2005.
- Saaty T., *Décider face à la complexité : une approche analytique multicritère d'aide à la décision*. Coll.

Pyroxène: un système d'aide à la décision territoriale par intégration de simulateurs spatiaux.

Application à l'évolution de la carte du risque d'incendie de forêt.

Eric Maillé* — Bernard Espinasse** — Sébastien Fournier**

**Cemagref, Unité de Recherche Ecosystèmes Méditerranéens et Risques
Département Gestion des Territoires, CS40061, Le Tholonet
13182 Aix en Provence cedex 5. Email : eric.maillé@cemagref.fr*

***Laboratoire des Sciences de l'Information et des Systèmes (LSIS)
UMR CNRS 6168, Université Paul Cézanne (Aix-Marseille III)
Domaine Universitaire de Saint-Jérôme, Avenue Escadrille Normandie-Niemen
13397 Marseille cedex 20. Email : bernard.espinasse@lsis.org, sebastien.fournier@lsis.org*

RESUME. La gestion à moyen terme de l'accroissement des risques territoriaux requiert des moyens de représentation et de simulation des dynamiques spatiales. Pour le cas du risque d'incendie de forêt, des modèles de dynamiques spatiales de zones combustibles et des zones vulnérables, à différentes échelles, existent. Pour assurer l'intégration des simulateurs de ces dynamiques de différentes thématiques et opérant à des niveaux d'échelle différents, nous proposons une infrastructure logicielle d'intégration à base d'agents spatiaux dans un environnement de système d'information géographique nommée Pyroxène.

ABSTRACT. Medium term management of territorial risk increase requires means to represent and simulate spatial dynamics. For forest fire risk, models of spatial dynamics of fuel zones and vulnerable zones exist. In order to integrate simulators of spatial dynamics of different thematic at different scale levels, we propose a spatial agent based and GIS based software infrastructure called Pyroxene.

MOTS-CLES : systèmes d'aide à la décision, systèmes multi agents, systèmes d'information géographique, couplage, intégration à base d'agents, dynamiques spatiales, risque d'incendie de forêt.

KEYWORDS: decision support systems, multi-agent systems, geographical information systems, coupling, agent based integration, spatial dynamics, forest fire risks.

1. Introduction

Les changements environnementaux actuels induisent l'émergence ou l'amplification de nombreux risques naturels tels que le risque d'inondation, d'incendie de forêt, etc. L'aide à la décision de gestion des risques environnementaux requiert la modélisation de l'interaction entre les dynamiques d'écosystèmes combustibles et la dynamique des systèmes anthropiques vulnérables. La représentation de « l'anthropo-écosystème » dans un modèle global unique se heurte tant à la complexité du système qu'à la distribution des connaissances nécessaires dans des disciplines très différentes. Par ailleurs, la décision de gestion territoriale est aussi multi-scalaire, puisqu'elle se réfère à des diagnostics et des champs d'intervention d'échelle spatiale différente.

Nous proposons une solution conceptuelle et opérationnelle d'un outil d'aide à la décision de gestion du risque lié à l'incendie de forêt fondée sur l'intégration de simulateurs de dynamiques spatiales existants, dans l'objectif de simuler l'évolution du risque lié à l'incendie de forêt au cours du temps. Ce système d'intégration présente deux caractéristiques majeures : il est multi-thématique, car il permet l'intégration de simulateurs qui implémentent des modèles se référant à des champs ontologiques différents : celui des écosystèmes, celui des dynamiques anthropiques et celui des risques naturels ; il est aussi multi-scalaire, en permettant l'interopérabilité de simulateurs dont les modèles représentent le système territorial à des niveaux d'échelle différents.

En section 2 de cet article nous proposons une présentation générale de la modélisation de la dynamique du risque lié à l'incendie de forêt par intégration multi-thématique et multi-scalaire de modèles disciplinaires. La section 3 est consacrée au cadre conceptuel d'un système d'intégration de modèles. Enfin, en section 4, nous présentons l'implémentation de ce cadre sous la forme d'une plateforme d'intégration de modèles à base d'agents appelée PYROXENE.

2. Multi-modélisation des dynamiques spatiales à l'origine de l'évolution du risque d'incendie par intégration de modèles

L'augmentation du risque d'incendie émane de la confrontation de deux dynamiques spatiales très différentes : la *dynamique des espaces combustibles*, principalement des espaces forestiers, et la *dynamique des espaces vulnérables*, en particulier des espaces « urbains discontinus » (Jappiot & al., 2000). Ces dynamiques sont contrôlées par un système anthropique de régulation, au travers de plans d'aménagement. La figure 1 explicite le lien entre les dynamiques spatiales et la dynamique du risque, ainsi que la limite du système modélisé.

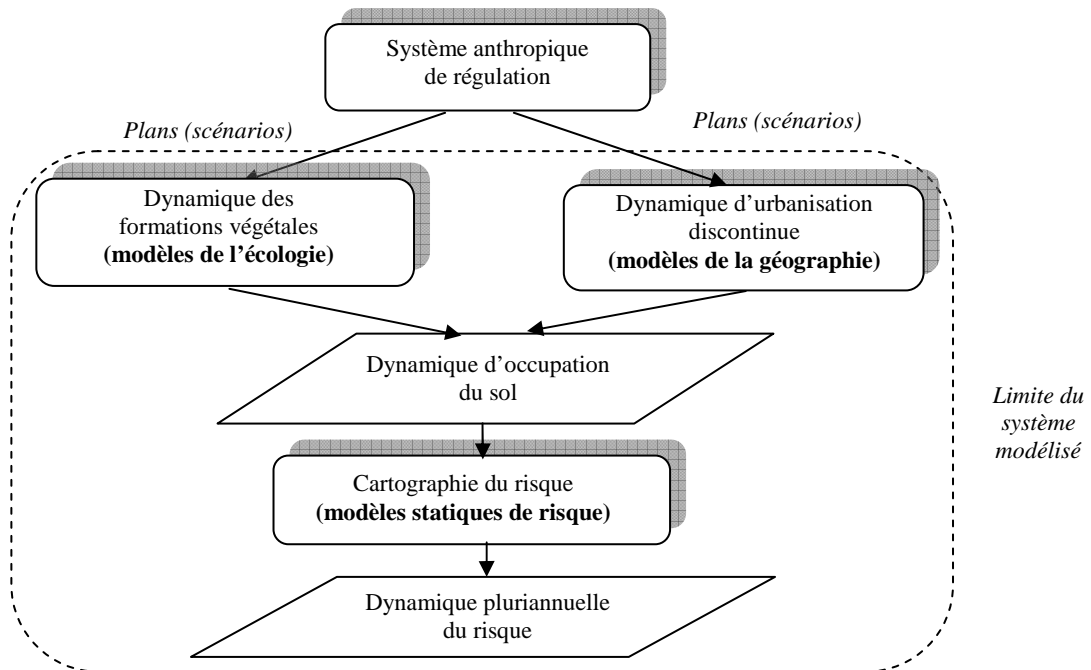


Figure 1. Intégration de modèles de dynamique d'occupation du sol pour la représentation de la dynamique du risque lié à l'incendie de forêt

Dans une perspective d'aide à la décision pour la planification de la gestion de l'espace, on ne cherche pas à modéliser le système anthropique de régulation, qui est en fait précisément l'utilisateur final (ou l'ensemble des utilisateurs finaux) auquel s'adresse l'outil envisagé. En revanche, cet utilisateur doit être en mesure de formuler et de tester des scénarios de plans d'aménagement, afin d'en évaluer l'impact sur l'évolution du niveau de risque tout au long de la durée du plan. Ce sont donc les processus sous-jacents à l'origine de cette évolution du risque qu'il convient de représenter.

2.1. Trois thématiques

Ce sont donc trois ensembles de connaissances thématiques qui doivent être formalisées en modèles opératoires puis intégrés pour la spécification d'un système de représentation spatialisée de la dynamique du risque : la thématique de la dynamique des écosystèmes, de la dynamique d'urbanisation discontinue et la thématique du risque d'incendie :

- la modélisation de la progression du combustible relève essentiellement de l'écologie qui fournit des modèles performants capables de représenter la dynamique des formations végétales, à différentes échelles. Certains modèles comme AFFORSIM (Prevosto & al., 2003) ou CAPSIS (de Coligny, 2006) tentent de représenter de façon individus-centrée le comportement spatial d'entités végétales (arbres). Ils se fondent sur les résultats de l'écologie fonctionnelle, décrivant en particulier les mécanismes de la reproduction des végétaux (dispersion des graines, compétition entre individus, etc.). Ils sont généralement appliqués à des extensions spatiales très limitées (de la parcelle au petit massif forestier). D'autres modèles, tels que SIERRA (Mouillot & al., 2001) ou LANDIS II (Scheller & al., 2006) considèrent plutôt le comportement global des formations végétales à l'échelle paysagère. Ils se fondent notamment sur les résultats de l'écologie du paysage, et sont applicables à des extensions spatiales plus grandes (du massif forestier à la région forestière).
- La modélisation de la dynamique des zones vulnérables, et de l'extension urbaine en particulier relève principalement de la géographie. Les modèles sont le plus souvent des automates cellulaires pour la représentation du phénomène de diffusion spatiale, modulés ou non par les caractères géographiques des espaces sur lesquels la diffusion se produit (Batty & al., 2000, Langlois & al., 1997, Dubos-Paillard & al., 2003, Ellerkamp, 2001).

Mais les travaux portant spécifiquement sur la modélisation du processus d'urbanisation discontinue, en zone périurbaine et à l'échelle des territoires locaux sont bien moins nombreux. Nous avons développé le modèle MACROPOLIS, qui est un automate cellulaire implémenté sur un système d'information géographique matriciel, spécialisé dans la représentation de la dynamique des espaces mixtes, associant les espaces naturels et les espaces urbains. Les fonctions de transition des classes d'automates sont fondées sur l'évaluation de la relation spatiale entre les différentes composantes des espaces mixtes, par le calcul d'indices de l'analyse spatiale matricielle.

Certains des processus anthropiques à l'origine de ces dynamiques peuvent aussi être représentés (Napoléone, 2005) et formalisés de façon analytique ou individus-centré : nous avons développé le système multi-agents MICROPOLIS (Maillé & al., 2007), dans lequel interagissent des agents sociaux, qui représentent des acteurs économiques du système territorial (propriétaires, acheteurs, aménageurs) et qui s'échangent des portions de l'espace (parcelles constructibles) afin de simuler le processus d'urbanisation discontinue en zone périurbaine.

- Enfin, L'évaluation du risque d'incendie en fonction de l'occupation de l'espace se fait grâce aux modèles cartographiques statiques de risque. Quatre grandes approches de modélisation du risque d'incendie de forêt peuvent être distinguées (Maillé & al., 2006) : la modélisation analytique attributaire permet le calcul d'un niveau de risque en chaque point de l'espace en fonction d'attributs géographiques de ce point (pente, couvert végétal, etc.) ; la modélisation spatiale agrégée évalue le niveau de risque en un point de l'espace en fonction de

l'occupation du sol de l'environnement (voisinage) de ce point ; la modélisation spatiale analytique évalue le risque en un point en fonction des aléas d'éclosion et des probabilités de propagation dans l'environnement de ce point ; enfin, la modélisation statistique est fondée sur l'étude statistique des résultats de simulation multiples de propagation des incendies. Toutes ces approches de modélisation conduisent à l'élaboration, pour un territoire donné, d'une carte du risque. Les modèles de risques sont toujours implantés sur Système d'Information Géographiques.

2.2. Deux niveaux d'échelles

Tant les dynamiques d'occupation du sol que la cartographie du risque d'incendie peuvent s'appréhender à des échelles diverses. Deux niveaux d'échelle sont pourtant privilégiés, dans une perspective d'aide à la décision pour la planification : l'aide à la décision pour la planification opérationnelle requiert l'intégration de modèles à grande échelle, pour laquelle sont appréhendés les entités géographiques élémentaires : parcelles, bâtis individuels, entités végétales (bosquet, voire arbre). L'extension sur laquelle opèrent les modèles va de la parcelle au quartier de commune, voire à tout l'espace communal. Les modèles de type individu-centrés sont souvent bien adaptés à ce niveau d'échelle. Nous appelons ce niveau d'échelle le niveau « micro » ; c'est aussi le niveau d'échelle auquel nous faisons référence par le terme « grande échelle ».

L'aide à la planification stratégique s'accommode d'échelles plus petites pour lesquelles les phénomènes peuvent être appréhendés par des modèles plus agrégés. Les modèles de diffusion à base d'automates cellulaires sont bien adaptés à ce niveau d'échelle. Concrètement, les entités géographiques élémentaires susceptibles d'être discriminées à ce niveau d'échelle sont la « zone agricole », le « petit massif forestier », la petite agglomération. L'extension sur laquelle les modèles sont mis en œuvre va de la commune jusqu'à un périmètre pluri-communal ou un grand massif forestier et ses abords. Nous appelons ce niveau d'échelle le niveau « macro » et c'est aussi le niveau d'échelle auquel nous faisons référence par le terme d'« échelle moyenne ».

L'articulation des deux échelles se situe à deux niveaux de perspectives différents :

- Au niveau de l'utilisation de l'outil d'aide à la décision il s'agit de contribuer à répondre à deux niveaux de questionnement différents : aide à la planification opérationnelle pour les modèles à grande échelle (micro), aide à la planification stratégique pour les modèles à échelle moyenne (macro). Dans cette perspective, il n'y a pas de réelle intégration des modèles.
- Au niveau du fonctionnement global du SADE, il s'agit pour les modèles micro de « localiser » les zones de la mise en œuvre du système des modèles « micro ». La « rencontre » des modèles à échelle moyenne permet la définition

des zones critiques ou des zones d'intérêt, à l'intérieur desquelles peuvent être initiées des sessions de simulation à l'échelle macro.

3. Le cadre conceptuel de l'intégration et la mise en cohérence des modèles.

Dans cette section, nous décrivons succinctement le cadre conceptuel général du système d'intégration ainsi qu'un cadre de spécification des modèles d'intégration assurant la mise en cohérence sémantiques des modèles à intégrer.

3.1. Le cadre conceptuel général

Nous disposons donc d'un total de 6 modèles, tous ayant une composante spatiale explicite, 4 étant des modèles dynamiques. Trois des six modèles s'appliquent aux grandes échelles, alors que les trois autres s'appliquent aux échelles moyennes. La production d'un outil d'aide à la décision intégré pour la gestion de l'évolution du risque d'incendie de forêt en lien avec la dynamique d'occupation de l'espace relève donc d'une problématique d'intégration de modèles fondée en premier lieu sur la spécification de différents types de **modèles d'intégration de modèles** (Maillé, 2008). La figure 2 récapitule le cadre de modélisation global.

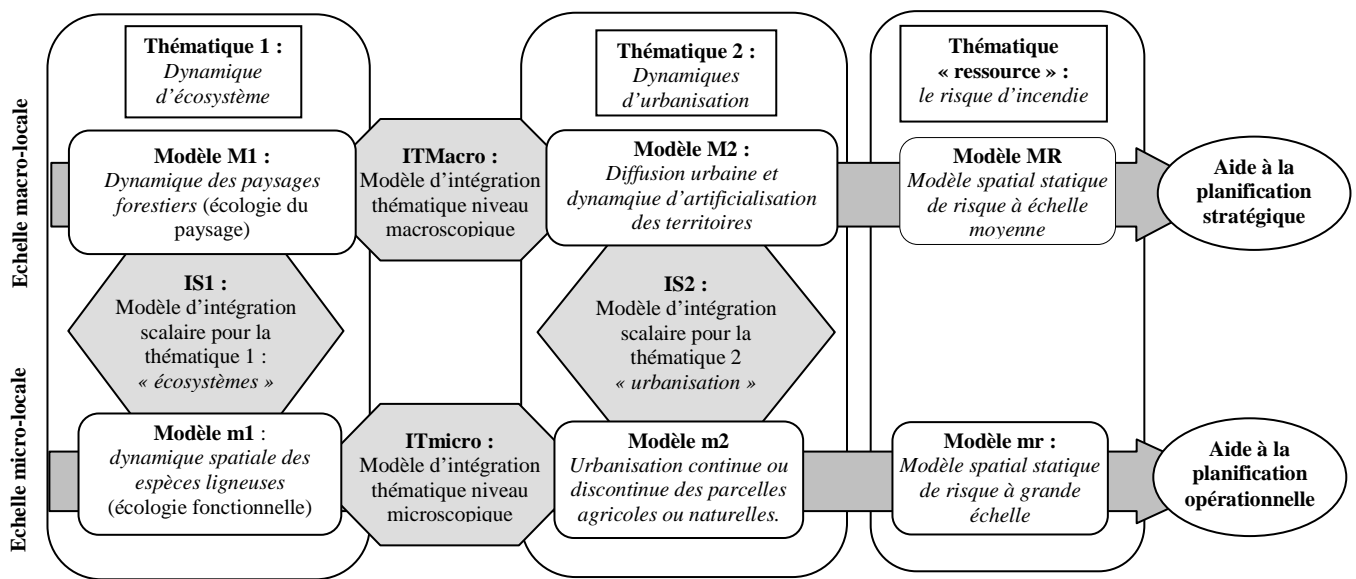


Figure 2. Cadre conceptuel de l'intégration : l'organisation des modèles d'intégration de modèles

Les modèles d'intégration de modèles ont pour but la mise en compatibilité sémantique de l'information produite par les différents modèles. Quatre modèles d'intégration de modèles doivent être spécifiés :

- deux modèles d'intégration thématique, soit un pour chaque niveau d'échelle (ITMacro et ITmicro).
- deux modèles d'intégration scalaire, soit un pour chaque thématique (IS1 et IS2).

Les modèles d'intégration de modèles doivent être spécifiés par l'utilisateur, en utilisant un cadre de spécification proposé par le système.

3.2. Le cadre conceptuel de spécification des modèles d'intégration

Un *cadre conceptuel de spécification* des modèles d'intégration est proposé (Maillé, 2008), organisé selon les trois vues du cadre conceptuel générique de représentation des systèmes spatio-temporels appelé triade de Peuquet, proposé par cet auteur en 1994 (Peuquet, 1994). Celui-ci propose la représentation du système selon trois vues : la vue sémantique, la vue spatiale et la vue temporelle. Par exemple, la vue sémantique d'un modèle d'intégration de deux modèles M1 et M2 peut être formalisée sous forme propositionnelle, comme illustré ci-après [1] :

$$\text{classe}CI_{\text{sémantique}SM1} \wedge_{M1} \text{classe}C2_{\text{sémantique}SM2} \rightarrow \text{classe}CI_{\text{sémantique}SM1} \quad [1]$$

Cette proposition s'interprète comme suit : lorsqu'une entité de classe C1 produite par le modèle M1 entre en interaction avec une entité de classe C2 produite par le modèle M2, une entité de classe CI est créée, appartenant à la sémantique (ensemble des classes) du modèle intégré. L'opérateur \wedge_{M1} , spécifique au modèle d'intégration, possède les propriétés de la conjonction. La vue spatiale permet de spécifier la dimension spatiale de l'interaction (en particulier les caractères spatiaux de l'entité générée en fonction des caractères spatiaux des entités initiales). La vue temporelle permet de spécifier la temporalité de l'interaction, en particulier l'impact de la chronologie des événements ayant affecté les entités initiales (notamment leur date d'instanciation) sur l'entité générée. Dans cet article, nous ne développons pas plus avant le cadre conceptuel de spécification des modèles d'intégration de modèles, qui fait l'objet de publications spécifiques (Maillé, 2008).

Mais outre ce niveau conceptuel, qui a pour but essentiel la levée des contraintes sémantiques de l'intégration, la spécification du système d'intégration recouvre aussi l'ensemble des éléments devant être spécifiés afin d'assurer l'interopérabilité des simulateurs en tant que systèmes logiciels, afin de permettre le développement d'un simulateur intégré composite mais cohérent, apte à répondre aux questions de l'utilisateur de façon efficace et fiable. Nous parlons, à ce niveau, d'intégration opérationnelle, dont l'objectif est la levée des contraintes syntaxiques de l'interaction entre les simulateurs, c'est-à-dire leur mise en interopérabilités, pour permettre le développement d'un outil opérationnel.

4. Implémentation, mise en œuvre et validation de la plateforme d'intégration Pyroxène.

Pyroxène est un prototype d'implémentation du système d'intégration thématique et scalaire de modèles de dynamique d'occupation du sol pour la simulation de l'évolution au cours du temps du risque d'incendie de forêt.

4.1. Principales caractéristiques générales du système d'intégration

Le système Pyroxène présente plusieurs caractéristiques majeures :

- Il s'agit d'un *système multi-agents d'intégration* développé en langage JAVA, sur la plateforme de développement de systèmes multi-agents cognitifs Tilab JADE, répondant à la norme FIPA (FIPA, 2001). L'intégration est réalisée au moyen d'agents adaptateurs, encapsulant les simulateurs spatiaux existants.
- Il s'agit d'un système distribué, partiellement architecturé selon la norme HLA (IEEE, 2000) autour d'une infrastructure de médiation centrale à base d'agents et de fédérés périphériques constitués par les simulateurs et les systèmes supports des fonctions de l'intégration.
- L'intégration se fonde sur l'exécution des modèles d'intégration thématiques et scalaires sur les systèmes d'information géographique matriciel GRASS ou vectoriel topologique ESRI ArcGIS ArcInfo Workstation.

L'implémentation de *Pyroxène* possède ainsi une structure figée permettant l'intégration des trois thématiques (écosystèmes, dynamiques urbaines et risque) et de deux niveaux d'échelle (micro-local et macro-local) (figure 3)

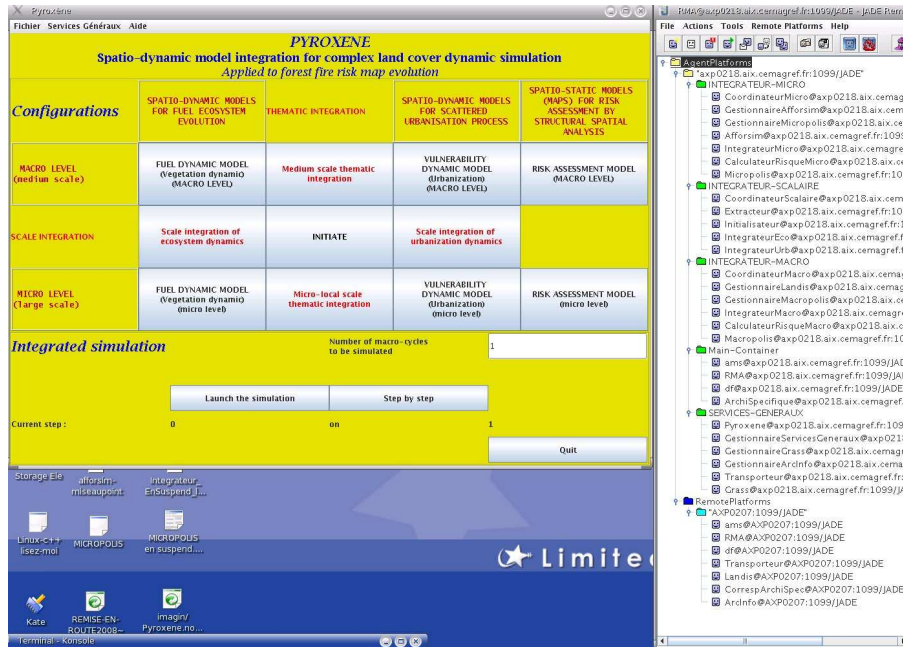


Figure 3. L'interface utilisateur de PYROXENE et modules du SMA d'intégration sous Tilab JADE

Pyroxene permet l'intégration de simulateurs de dynamiques spatialisées répondant à deux contraintes principales :

- ils doivent admettre en entrée et produire en sortie des couches d'information géographique « sémantiques », matricielles (classes) ou vectorielle.
- Ils doivent admettre en paramètre le nombre de cycles à simuler.

4.2. Elaboration des scénarios

L'élaboration de scénarios par l'utilisateur consiste en la définition de deux ensembles de données (figure 4) :

- les plans d'aménagements fournis à chacun des modèles sous la forme d'information géographique pour « contrôler le déroulement de la simulation ». Y sont spécifiés, par exemple, des zones d'exclusion pour l'installation de la forêt ou du bâti.
- Les valeurs de paramètres spécifiques à chacun des modèles, représentant l'environnement général (écologique, climatique, économique, etc.) de la simulation

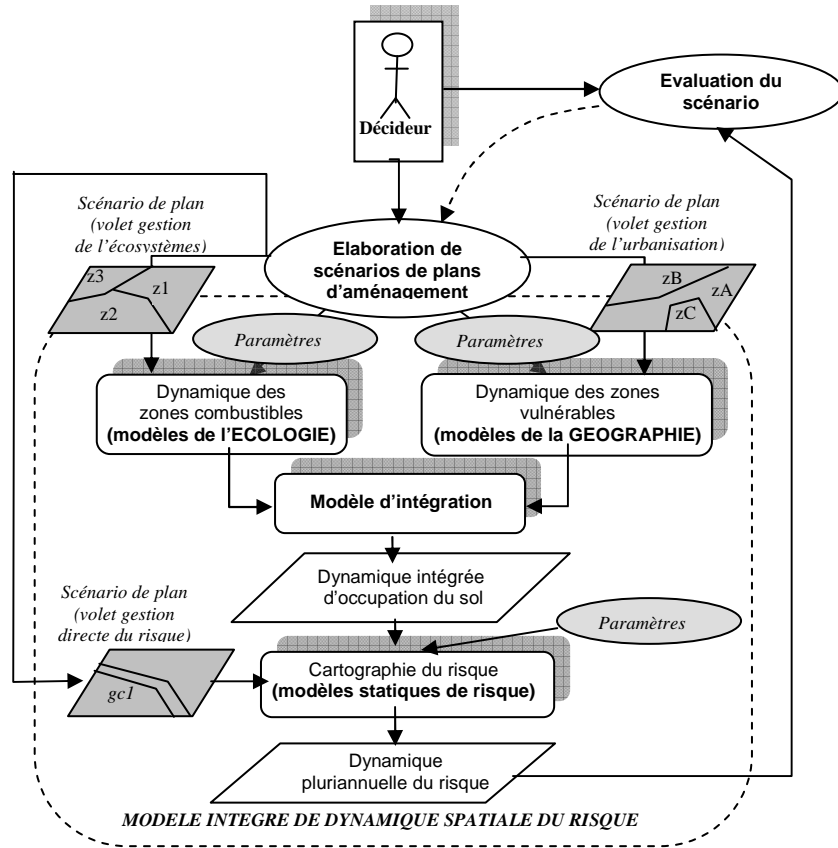


Figure 4. Cycles « élaboration des scénarios, simulation intégrée, évaluation des scénarios ».

Les plans d'aménagement élaborés dans le cadre des scénarios peuvent avoir pour objectif soit l'infléchissement des dynamiques spatiales, soit la limitation directe du risque : ils modifient les données d'entrée initiales des modèles thématiques. Par définition, chacune de leurs modalités sont datées (par exemple, installation d'une coupe de combustible *gc1* à compter de la date *d*).

Les paramètres sont eux strictement spécifiques de chacun des modèles initiaux et sont renseignés selon les modalités requises par leur simulateur respectif.

4.3. Fonctionnement de l'architecture à base d'agents

Outre le choix et la configuration des modèles de simulation, Pyroxene permet la gestion des simulations intégrées mais également la spécification par l'utilisateur des *modèles d'intégration* scalaire ou thématique.

Lors du lancement de la simulation intégrée, les divers agents échangent leurs messages afin de coordonner l'exécution par les différents systèmes logiciels des différentes tâches requises : simulations (par les simulateurs), exécution des modèles d'intégration (par les SIG), mise en forme syntaxique et échanges (transport) de données (par divers applicatifs ou protocoles), etc (figure 5).

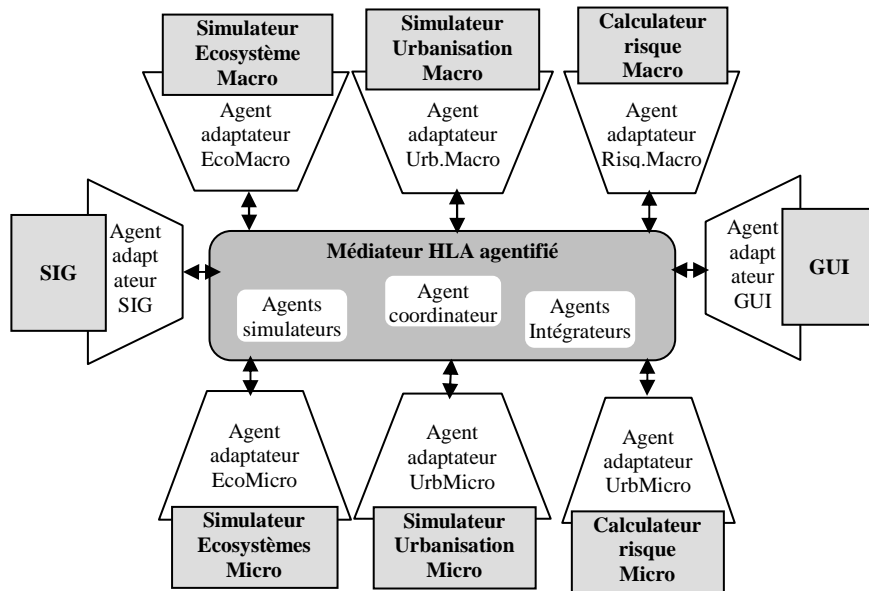


Figure 5. Architecture basée HLA du système multi-agents d'intégration

L'exécution au niveau du système d'information géographique des modèles d'intégration de modèles est opérée sous le contrôle d'agents *Intégrateur* (un agent intégrateur par thématique et par niveau d'échelle).

4.4. Mise en œuvre et validation

Pyroxene est actuellement mis en œuvre sur une zone expérimentale de la conurbation d'Aix-Marseille, regroupant 43 communes de la Communauté d'Agglomération du Pays d'Aix.

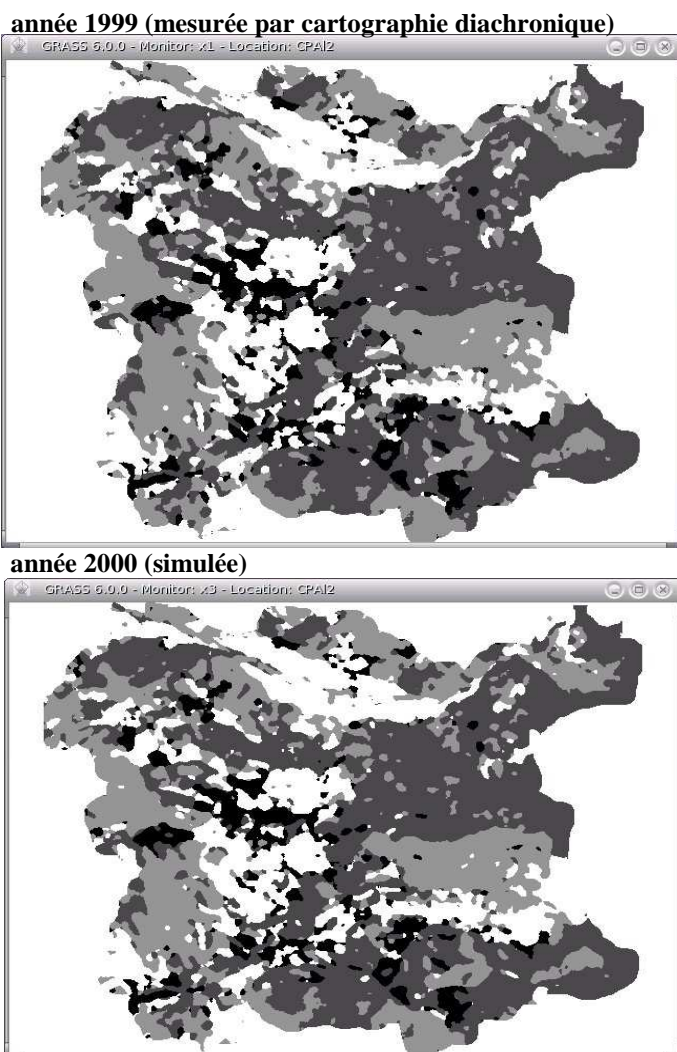
Dans la version actuelle du système, ce sont principalement les simulateurs de dynamique d'écosystème Afforsim (Prévosto & al., 2003), pour l'échelle micro-locale, et Landis II (Scheller & al., 2006), pour l'échelle macro-locale qui ont été testés. Les simulateurs des dynamiques d'urbanisation discontinue testés sont Micropolis et Macropolis (Maillé, 2008).

Les protocoles de validation de la dynamique du niveau de risque étant encore en cours de mise au point par les cindyniciens, la validation porte essentiellement sur la qualité de la représentation intégrée de la dynamique d'occupation du sol. La validation s'appuie sur la comparaison des simulations intégrées des dynamiques passées avec les résultats de travaux de cartographie diachronique à partir d'images télé-acquises (satellites et aériennes) (Maillé & al., 2005).

Il existe deux niveaux de validation : la validation des modèles initiaux, par thématique et par niveau d'échelle, puis la validation des simulations intégrées. Notre objectif actuel est essentiellement l'estimation de l'erreur issue de l'intégration elle-même, en éludant l'erreur des modèles initiaux.

Globalement, si les résultats simulés semblent cohérents avec les résultats expérimentaux d'un point de vue quantitatif agrégé, des écarts beaucoup plus importants sont révélés pour ce qui concerne les distributions spatiales. En particulier, les contours des zones d'extension non contraintes (urbaines ou forestières) semblent difficilement prédictibles. L'outil opérationnel semble donc utilisable dans une perspective d'aide à la décision stratégique agrégée, mais ne pourra probablement pas constituer l'unique outil d'aide à la décision de planification opérationnelle localisée.

Au travers de modèles de risque, Pyroxene permet finalement la représentation de l'évolution de la carte des niveaux de risque en fonction de la dynamique d'occupation du sol simulée (figure 6).



Légende :

Risque faible ou nul Risque modéré Risque moyen Risque fort

Figure 6. Exemple de représentation de l'évolution des niveaux de risque par simulation intégrée

5. Conclusion

La gestion territoriale en général, et la gestion du risque en particulier, se fondent sur des connaissances multi-sectorielles, se rapportant le plus souvent à plusieurs niveaux d'échelles. L'intégration formelle de telles connaissances au sein des systèmes d'information a pour objectif d'assurer la cohérence de l'aide à la décision territoriale fournie par ces systèmes. Dans une perspective dynamique, et en particulier pour la planification territoriale, l'intégration de modèles dynamiques permet la représentation de l'interaction entre des processus représentés par des champs disciplinaires différents.

Nous proposons donc un système d'intégration de simulateurs de dynamique d'occupation du sol et de modèles statiques de risque capable de prendre en charge la mise en cohérence sémantique de modèles issus de champs disciplinaires variés et l'interopérabilité syntaxique de leur simulateur implémenté. Ce système présente la particularité de se fonder sur le paradigme *agents* en valorisant les capacités de traitement de l'information sémantiques des agents, ainsi que leurs capacités interactionnelles essentielles dans les démarches d'interaction.

La validation d'un tel système est une perspective de long terme, nécessitant tout d'abord la validation des modèles initiaux, puis la validation du modèle intégré de dynamique d'occupation du sol, et enfin la validation du modèle dynamique de risque. Ce travail doit s'appuyer sur une démarche expérimentale lourde associant les spécialistes des différentes disciplines mises en jeu et les informaticiens.

D'un point de vue opérationnel, le système n'a pas d'ambition prédictive, mais seulement prospective, en permettant aux aménageurs de comparer l'impact sur l'évolution du risque de différents scénarios de contrôle des dynamiques spatiales et d'aménagement du territoire. Il ne constitue pas l'unique outil d'aide à la décision pour la planification territoriale, mais doit être lui-même intégré dans une panoplie d'outils plus large, incluant des systèmes numériques opératoires aussi bien que des outils cartographiques et textuels traditionnels susceptibles d'intervenir dans l'interprétation globale des résultats.

6. Bibliographie

- Batty M. and Jiang B., « Multi-agent Simulation: Computational Dynamics within GIS », Innovation in GIS VII: Geocomputation, Martin D. and Atkinson P., Taylor & Francis, 2000, p. 55-71
- De Coligny F., "Efficient Building of Forestry Modelling Software with the Capsis Methodology.", In: Proceedings of Plant Growth Modeling and Applications PMA06, IEEE Computer Society, 2006, Fourcaud T, Zhang XP, eds., Los Alamitos, California: pp. 216-222
- Dubos-Paillard E., Guermont Y., Langlois P., « Analyse de l'évolution urbaine par automate cellulaire », 2003, UMR 6063 IDEES, Laboratoire MTG Rouen
- Ellerkamp P., Etude de l'extension urbaine autour d'Avignon et de Montpellier de 1936 à 1999 par la simulation de dynamiques spatiales à l'aide de modèles cellulaires d'interaction locale, Thèse de Doctorat, Université d'Avignon, 2001.
- FIPA (Foundation for Intelligent Physical Agents), Agent Software Integration Specification, Doc. number XC00079B, 2001.
- IEEE Institute of Electrical and Electronics Engineers, 1516-2000, Standard for Modeling and Simulation (MetS), High Level Architecture (HLA), 2000.
- Jappiot M., Blanchi R., Alexandrian D., « Cartographie du risque d'incendie de forêt : besoins, méthodes et données, essai de normalisation, *Actes des Journées GIS Incendie de Forêt*, Octobre 2000, Hyères.
- Langlois A., Phipps M., « Automates cellulaires. Application à la simulation urbaine. » Hermes. Paris, 1997.
- Maillé E., Bouillon C., Consommation des espaces agricoles et naturels par le processus d'urbanisation discontinue, Rapport de Recherche, Cemagref, Aix-en-Provence, 2005.
- Maillé E., Espinasse B., Un cadre de modélisation pour la simulation de dynamiques spatiales complexes, in. Actes du Colloque SAGEO07, Rencontres Internationales de Géomatique, 18-19 juin 2007, Clermont-Ferrand.
- Maillé E., Intégration conceptuelle et opérationnelle de modèles spatiaux-dynamique, Thèse de Doctorat, Université Paul Cézanne Aix-Marseille III, 2008.
- Mouillot F., Rambal S., Lavorel S., A generic process-based simulator for Mediterranean landscapes (SIERRA): Design and validation exercises., in. *Forest Ecology and Management* 147: 75-97. 2001.
- Napoléone C., « Prix fonciers et immobiliers, et localisation des ménages au sein d'une agglomération, urbaine », thèse de Doctorat, Ecole des Hautes Etudes en Sciences Sociales (EHESS), GREQUAM, 2005.
- Peuquet D. J., "It' About Time: A Conceptual Framework for the Representation of Temporal Dynamics in Geographic Information System." *Annals of the Association of American Geographers* 84(3), 1994, Blackwell Publishers, pp.441-461

Urban versus Rural: the decrease of agricultural areas and the development of urban zones analyzed with spatial statistics

Beniamino Murgante — Giuseppe Las Casas — Maria Danese

Laboratory of Urban and Territorial Systems - University of Basilicata
10 Via dell'Ateneo Lucano
85100 Potenza – Italy

Beniamino.murgante@unibas.it, maria.danese@unibas.it

ABSTRACT. The expression “periurban area”, despite its large use, does not have a clear and unambiguous definition. Different points of view generate more barriers in determining the exact edge of these zones. These various approaches are due to the complexity of the phenomenon to be analyzed and to the huge variety of territorial contexts in which it may reveal. The phenomenon is characterized by an urban growth with soil consumption generating loss of competitiveness for agricultural activities. Some experiences, in a spontaneous way, take into account only proximity to urban areas. It is obvious that contiguity condition alone is not sufficient to define such a complex phenomenon. The aim of this paper is to define more precise rules in order to describe the periurban phenomenon, using techniques of spatial statistic and point pattern analysis. This approach has been tested in the case of study of Potenza Municipality. This area, about 18,000 ha wide, is located about 900 m a.s.l. and is characterized by a relevant presence of forests, very steep slopes and a large number of areas with geological instability phenomena. Our interest in this area especially comes after the earthquake of 1980, when a huge migration of inhabitants began towards the countryside around the Town of Potenza. It was so intense that it deeply altered the urban morphology and modified the traditional physical functional relationships between the town and the rural territory.

KEYWORDS: Spatial statistics, Autocorrelation, Periurban areas, Urban sprawl, soil consumption, land use planning .

1. Introduction

Study of rural areas has generally very few interest from researchers, if compared to a greater attention to urban or metropolitans areas, so it pays a great price in terms of very scarce knowledge of its territory.

Only in recent times the expression “periurban area” has been frequently used in planning documents. In planning literature, the periurban phenomenon has been defined in different ways; Thomas Le Jeannic (1997) describes the population displacement as the need to escape from the dense city in order to have more space and a better environment. Also, growth of periurban belt is due to high costs of flats, the need of individual dwellings and land rents (Guerois and Pumain 2001). Since 1980s it is less and less possible to distinguish town from country, denying the concept of two separate entities which was for many years one of the cornerstones of spatial planning (Hidding, et al. 2000; Van Den Berg and Wintjes 2000). Rural areas are more urbanized and periurban belt have increased the number of inhabitants with an uncontrolled growth. At the same time, urban areas have lost resident population gaining population in transit (Alberti, et al. 1994) because of the activities concentration in urban areas. All these situations produce a huge commuting phenomenon (Cavailhès, et al. 2004). The main feature of this trend is a low density of urbanization which spreads in all directions (Camagni, et al. 1998). Growth of these areas is strictly related to urban sprawl, generating negative repercussions to agricultural activities. A great amount of roads have been built to improve dwellings accessibility and car is the only means of transport (Camagni, et al. 2002).

This is an opposite tendency compared to the period after the Second World War, when urban planners used statistical methods to give a dimension of the migratory phenomenon towards towns. Urban sprawl is so complex to analyze, that classical statistics are not enough for a complete understanding of the phenomenon. Settlement location in zones surrounding urban areas takes into account environmental features, accessibility, agricultural losses of productivity. In order to achieve a more complete analysis it is important to analyze each phenomenon with its spatial location, so that it is possible to consider the concentration of some events in some areas and their possible interactions. Geostatistics can be useful in order to study this problem with an innovative approach compared to the classic socio-economic techniques. This method allows an analysis which may determine the actual trend in one region. This technique has been applied in Potenza Municipality, where a migratory phenomenon began from urban to rural areas after a strong earthquake occurred in 1980.

All the informative layers have been combined with a land suitability procedure in order to define a periurban fringe with a certain precision.

2. An overview of spatial statistics techniques

The main aim of spatial analysis is a better understanding of spatial phenomena aggregations and their spatial relationship. Spatial statistical analyses are techniques which use statistical methods in order to determine if data show the same behaviour of the statistical model. Data are treated as random variables. The *events* are spatial

occurrences of the considered phenomenon, while *points* are each other arbitrary locations. Each event has a set of attributes describing the nature of the event. *Intensity* and *weight* are the most important attributes; the first is a measure identifying the event strength, the second is defined by the analyst who assigns a parameter in order to define if an event is more or less important according to some criteria. Spatial statistics techniques can be grouped in three main categories: *Point Pattern Analysis*, *Spatially Continuous Data Analysis* and *Area Data Analysis*.

The first group considers the distribution of point data in the space. They can follow three different criteria:

- random distribution: the position of each point is independent of the others points;
- regular distribution: points have an uniform spatial distribution;
- clustered distribution: points are concentrated in some building clusters.

The second group takes into account the spatial location and the attributes associated to points, which represent discrete measures of a continuous phenomenon. The third group analyzes aggregated data which can vary continuously through space and can be represented as point locations. This analysis aims to identify the relationships among variables and *spatial autocorrelation*. If some clusters are found in some regions and a positive spatial autocorrelation is verified during the analysis, it can describe an attraction among points. The case of negative spatial autocorrelation happens when deep differences exist in their properties, despite the closeness among events. It is impossible to define clusters of the same property in some areas; a sort of repulsion occurs. Null autocorrelation arises when no effects are surveyed in locations and properties. Null autocorrelation can be defined as the case in which events have a random distribution over the study area (O'Sullivan and Unwin, 2002). Essentially, the autocorrelation concept is complementary to independence: events of a distribution can be independent if any kind of spatial relationship exists among them.

Spatial distribution can be affected by two factors:

- first order effect, when it depends on the number of events located in one region;
- second order effect, when it depends on the interaction among events.

If these two definitions seem more clear, it isn't as much clear as the recognition of these effects over the space.

2.1. Kernel density

Kernel density is one of the point pattern analysis techniques, where input data are point themes and outputs are grids. While simple density computes the number of events included in a cell grid considering intensity as an attribute, kernel density takes into account a mobile three-dimensional surface which visits each point. The output grid classifies the event S_i according to its distance from the point S , which is the centre of the ellipse generated from the intersection between the surface and the plane containing the events (Bailey and Gatrell, 1995). The influence function defines the influence of a point on its neighbourhood. The sum of the influence functions of each point can be calculated by means of the density function, defined by:

$$\lambda(L) = \sum_{i=1}^n \frac{1}{\tau^2} k\left(\frac{L-L_i}{\tau}\right) \quad [1]$$

where:

- λ is the distribution intensity of points, measured in L ;
- L_i is the event i ;
- K is the kernel function;
- τ is the bandwidth.

The first factor influencing density values is bandwidth: if τ is too big the value of λ is closer to simple density; if τ is too small the surface does not capture the phenomenon. The second factor influencing density values is cell size like in every grid analysis.

2.2. Straight Line Distance

The straight line distance is a function measuring the distance between each cell and the nearer source. The source can be in vector or grid format. In the case of grid format some cells will contain information about the source and some others will not, while in the case of a vector theme it will be necessary a previous transformation in grid before determining the distance.

The output of straight line distance is in grid format and the distance is measured between the barycentres of the cells. Also, in this case it is important to estimate some factors such as the maximum distance within which one has to assess measures and sizes of cells.

2.3. Moran index

Moran index (Moran, 1948) allows to transform a simple correlation into a spatial one. This index takes into account the number of events occurring in a certain zone and their intensity. It is a measure of the first order property and can be defined by the following equation:

$$I = \frac{N \sum_i \sum_j w_{ij} (X_i - \bar{X})(X_j - \bar{X})}{(\sum_i \sum_j w_{ij}) \sum_i (X_i - \bar{X})^2} \quad [2]$$

where:

- where: N is the number of events;
- X_i and X_j are intensity values in the points i and j (with $i \neq j$), respectively;
- \bar{X} is the average of variables;
- $\sum_i \sum_j w_{ij} (X_i - \bar{X})(X_j - \bar{X})$ is the covariance multiplied by an element of the weight matrix. If X_i and X_j are both upper or lower relative to the mean, this term will be positive, if the two terms are in opposite positions compared to the mean the product will be negative;
- w_{ij} is an element of the weight matrix which depends on the contiguity of events. This matrix is strictly connected to the adjacency matrix.

There are two methods to determine w_{ij} ,: the “Inverse Distance” and the “Fixed Distance Band”. In the first method, weights vary in inverse relation to the distance among events:

$$w_{ij} = d_{ij}^{-z}$$

where z is a number smaller than 0.

The second method defines a critical distance beyond which two events will never be adjacent. If the areas to which i and j belong are contiguous, w_{ij} will be equal to 1, otherwise w_{ij} will be equal to 0. *Moran index* I can have values included between -1 and 1.

If the term is high, autocorrelation is positive, otherwise it is negative. Moran index vanishes in very rare cases, but usually the convergence is towards the theoretical mean value $E(I)$, where each value is independent from the others.

$$E(I) = -\frac{1}{N-1}$$

If $I < E(I)$, the autocorrelation is negative, if $I > E(I)$ the autocorrelation is positive.

The significance of Moran index can be evaluated by means of a standardized variable $z(I)$ defined as:

$$z(I) = \frac{I - E(I)}{S_{E(I)}}$$

where $S_{E(I)}$ is the standards deviation from the theoretical mean value $E(I)$.

2.4. Local Indicator of Spatial Association: the G function by Getis and Ord

Both LISA and G function take into account disaggregated measures of autocorrelation, considering the similitude or the difference of some zones. These indexes measure the number of events with homogenous features included within a distance d , located for each distribution event. This distance represents the extension within which clusters are produced for particularly high or low intensity values.

The Local Indicator of Spatial Association (Anselin, 1995) is defined as:

$$I_i = \frac{(X_i - \bar{X})}{S_x^2} \sum_{j=1}^N (w_{ij} (X_j - \bar{X})) \quad [3]$$

where symbols are the same used in Moran's I, except for S_x^2 which is the variance.

The function by Getis & Ord (1992) is represented by the following equation:

$$G_i(d) = \frac{\sum_{i=1}^n w_i(d) x_i - \bar{x}_i \sum_{i=1}^n w_i(d)}{S(i) \sqrt{\left[(N-1) \sum_{i=1}^n w_i(d) - \left(\sum_{i=1}^n w_i(d) \right)^2 \right] / N - 2}} \quad [4]$$

which is very similar to Moran index, except for $w_{ij}(d)$ which, in this case, represents a weight which varies according to distance.

3. The case study

These techniques have been applied in Potenza municipality. It is located in the southern Apennines area of Italy and it is the chief town of Basilicata, which is a region with a very low residential density.

Its territory, about 18000 hectares sized, lays on an area roughly 900 m a.s.l., characterized by a relevant presence of forests, by very steep slopes and by a large number of areas interested by landslides. Although the morphology of this area did not encourage the diffusion of isolated houses, existing houses in the extra urban territory increased by 25% during the 1980-1990 decade. The migration of people from urban areas to rural areas happened after the earthquake in 1980, when a huge number of houses in the urban centre were defined uninhabitable and the fear of new seismic events, mixed with the wish of running away from the most damaged places, made citizens choose new localizations to build new houses. All this, together with other causes, such as social, economic and cultural ones, involved an enormous insediative process in the countryside around Potenza town, so intense that it deeply altered the urban morphology and modified the traditional physical functional relationships between the town and the extra urban territory.

This process caused a decrease of agricultural areas and a development of new settlements with a crown shape all around the town, with no respect for natural morphology and without considering the consequent resources consumption.

This kind of uncontrolled expansion went on uncontrolled also in the seventeen years after the 1980 earthquake and caused in territory the following phenomena, present nowadays:

- Diffuse territorial decline, with serious hydrogeologic accident phenomena deriving from anthropic uses (construction of small road infrastructures);
- Considerable abandonment of potentially agricultural soils related to expectations of estate development, especially in areas close to the town;
- Scattering of the extra urban insediative system, with inadequacy of facilities and infrastructures;
- Pauperization of environmental and landscape valuable components.

This situation emphasized the need of an urbanistic tool in order to control building in rural areas, with a strategic value for the whole municipal territory.

On this purpose, it can be useful to apply spatial statistic techniques to define criteria concerning suitability in building new settlements, considering particular tendencies that some specific areas already manifest.

A first factor considered in this study is density. Periurban area is characterized by a spread of settlements with extensive features, compared to the urban area which

has a greater density. Lower density is the first condition distinguishing periurban areas from urban ones.

Rural sites have a strong connection with agricultural activities and the relationship with the urban area is weak. It is also necessary to establish a lower threshold which can distinguish periurban areas from rural ones.

In order to calculate density, all the polygons representing buildings have been converted in points which are the events to take into account in point pattern analysis. The ratio between the number of flats and the number of buildings has been calculated from census data; this value has been considered as the intensity of events. Figure 1 compares the density of scattered settlements between 1987 and 2004 and it shows the huge growth of urban sprawl.

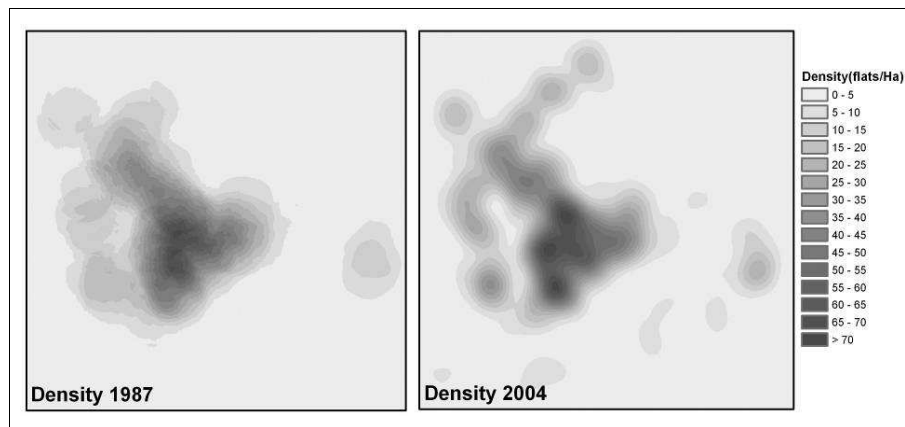


Figure 1. Density of scattered settlements in 1987 and 2004 (flats/hectare).

In the case of study, a value of bandwidth of 400 m and a cell size of the grid of 10 m have been used. A first rough analysis of periurban fringe takes into account zones with a low density expansion including areas with values of kernel density included between 1 and 18 flats/ha (fig.2). Orography and accessibility define the second factor, which consists of the distance from infrastructures because urban growth is more concentrated along the main line of road network. In order to locate areas with a good accessibility, distance from infrastructures has been defined so that it represents the tendency. Straight Line Distance identified areas with a distance from the main infrastructures within 200 meters. The third factor is the spatial autocorrelation which has been analyzed considering Moran Index, G function as developed by Getis & Ord and Local Indicator of Spatial Association (LISA).

In this case intensity of events is obtained as the ratio between number of inhabitants and number of buildings in each census zone. Moran index is able to specify if an event is clustered, scattered or with a random distribution. It has been

calculated by means of the inverse distance method considering data in two different periods, 1987 and 2004, to evaluate the variation of scattered rate of settlements. The following values have been achieved:

- Moran Index at 1987: $I_{1987} = 0.0698$;
- Moran Index at 2004: $I_{2004} = 0.0722$.

These two indexes show a low autocorrelation in both cases, and the second one is higher than the first one. These data can be interpreted as growth of settlements concentrated in some particular zones.

The next step of our study was to calculate the contiguity belt considered as the area where the phenomenon grows homogeneously and where it will intensify in the future. Moran index depends from the distances among points; it is possible to calculate a distance value which produces an index I with the maximum level of correlation among events, by maximizing the deviation z . This value has been calculated in 1600 m and it has been used as an input parameter in LISA and then in Getis & Ord function determining zones where events are autocorrelated.

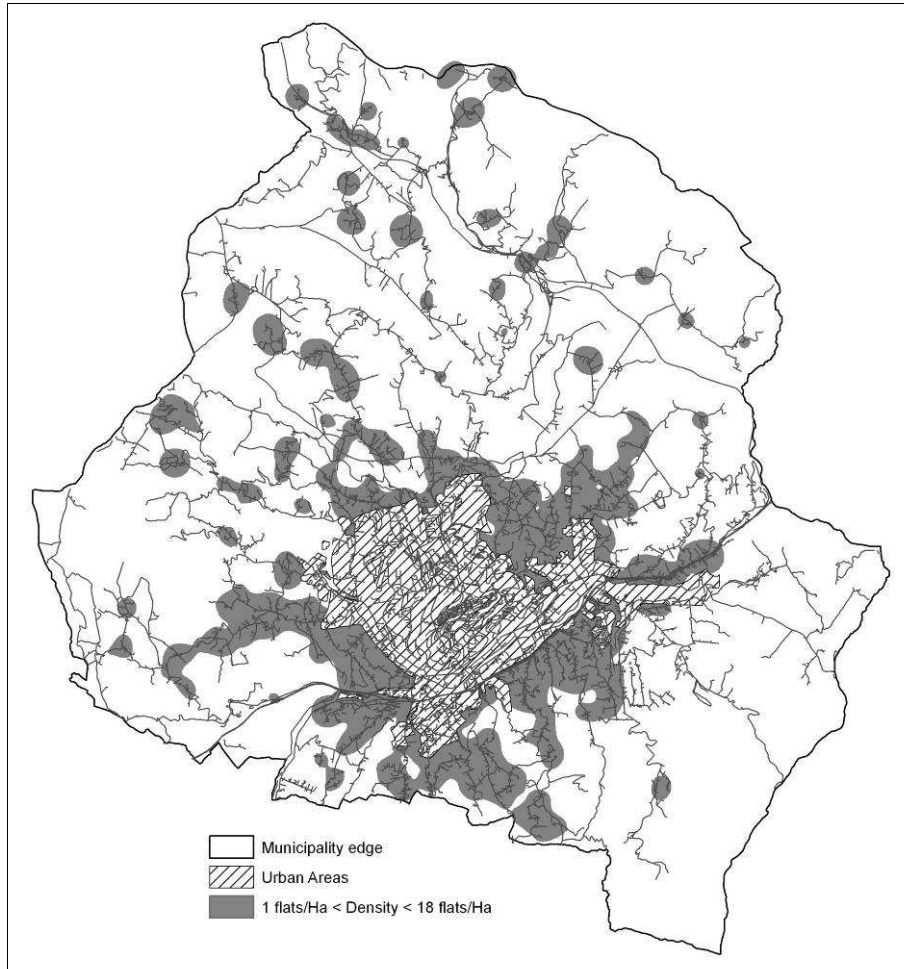


Figure 2. Areas which have Kernel density included between 1 and 18 flats/hectare.

A LISA positive value indicates a positive autocorrelation; obviously a negative autocorrelation corresponds to a negative value.

For the periurban fringe it is important to pay attention to the medium-low level of intensity, so the classes in table 1 have been considered.

Class	Autocorrelation	LISA
no correlation	Negative autocorrelation	-106,9 ÷ 0
1 low	Positive autocorrelation among lower bounds	0 ÷ 14
2 low	Positive autocorrelation among low bounds	14 ÷ 28
3 low	Positive autocorrelation among medium-low bounds	28 ÷ 54
4 high	Positive autocorrelation among high bounds	54 ÷ 84.7

Table 1. *LISA classes.*

In Getis & Ord function, highest and lowest values of G mean highest and lowest values of phenomenon intensity.

The classes in table 2 have been considered.

Class	Autocorrelation	Intensity value X_i (inhabitants/buildings)	G*
no correlation	Negative autocorrelation	$X_i \leq 18$ $X_i \geq 18$	-1.3 ÷ - 2 -6.3 ÷ 1
1 low	Positive autocorrelation among lower bounds	$X_i \leq 18$	-1.3 ÷ - 2
2 low	Positive autocorrelation among low bounds	$X_i \leq 18$	-2 ÷ - 4
3 low	Positive autocorrelation among medium-low bounds	$X_i \leq 18$	-4 ÷ - 6.3
4 high	Positive autocorrelation among high bounds	$X_i \geq 18$	1 ÷ 11.9

Table 2. *Getis & Ord function classes.*

Next figure (Fig. 3) compares results, showing the similitude of areas with positive autocorrelation achieved with both indicators. In our study case Getis & Ord function fits the phenomenon better because it is more sensible to intensity changes allowing more accurate classification.

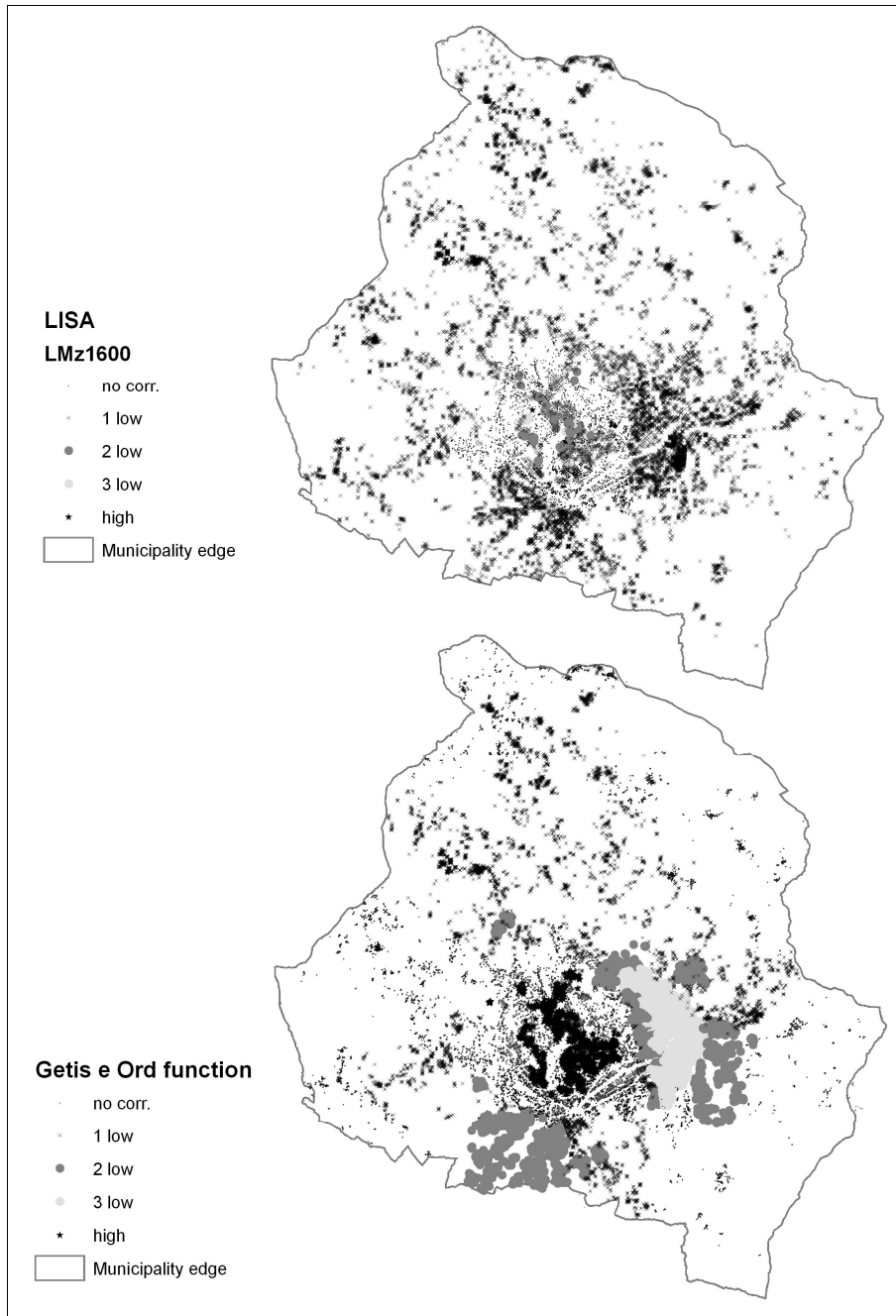


Figure 3. *Clusters location.*

Figure 4 shows in a more detailed map how deep are analysis produced with Getis & Ord function. Opposite values can occur in contiguous areas. Picture a shows how the highest values of autocorrelation correspond to the highest buildings of the town. Picture b highlights an abrupt transition, in a few metres, from no correlation to high correlation passing from ancient low buildings to high concrete buildings. Picture c shows how very elevated values of autocorrelation correspond to high buildings separate by narrow streets.

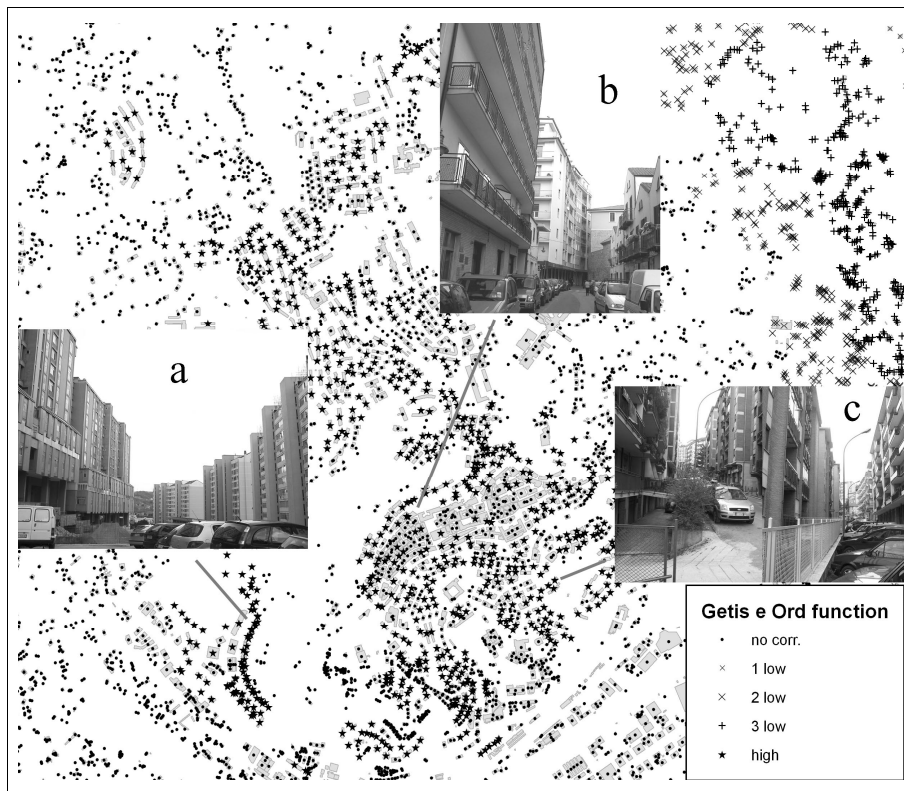


Figure 4. Detailed map with pictures which show autocorrelation difference.

4. Results and final discussion

Autocorrelation phenomena included in medium medium-low values have been interpolated thus generating polygons which represent the contiguity belt. These polygons represent the second level of suitability. It is composed by the inclusion rules considered in land suitability procedures reduced considering the global and local measures of autocorrelation. It is obvious that kernel density (Fig. 1 and 2) is a rough measure which needs a deeper analysis. Moran index and Getis & Ord

function give a further interpretation of phenomena considering contiguity not in all directions but only in some zones.

The exclusion rules in figure 5 have been considered in the present study: areas included within a distance of 150 m from rivers, streams and springs, slopes higher than 35%, Nature 2000 sites, hydro-geological risk zones, areas higher than 1200 m a.s.l., landslides, areas close to railways and road networks.

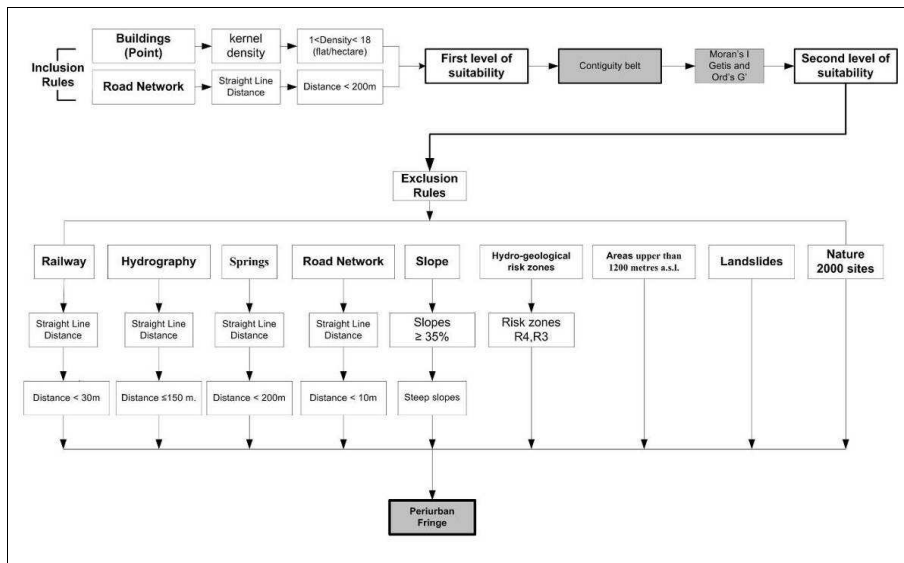


Figure 5. Scheme of the land suitability procedure for the location of Peri-urban fringe.

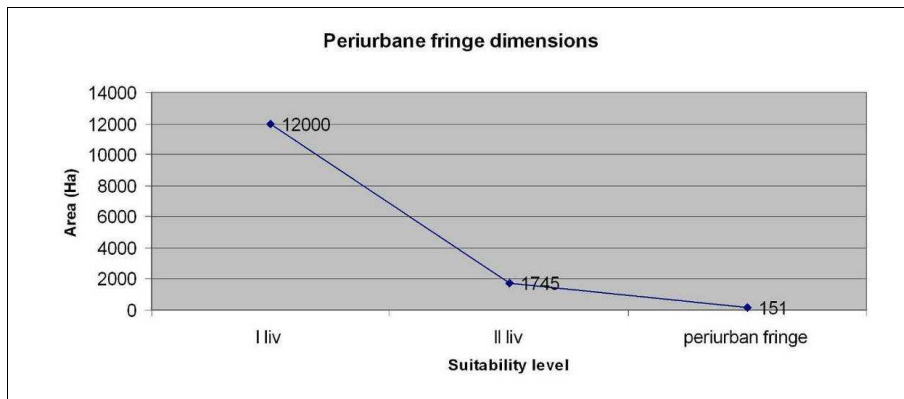


Figure 6. Size of suitable areas.

Figure 5 shows the flow chart of the land suitability procedure for the location of Peri-urban fringe. All these rules have been combined using map algebra techniques. Figure 6 quantifies the reduction of suitable areas achieved after the procedure.

The results are illustrated as geographic components in figure 7. Location of the contiguity belt is determined by the highway, which determines two gates for the town. In these areas urban sprawl is more intensified, particularly in the eastern part where the other road, which connect the industrial areas of Potenza with FIAT factory, amplifies the phenomenon. Steep slope obstructs urban growth in other zones. Periurban fringe (oblique hatch in figure 7) considers contiguity belt after the exclusion rules and represents areas suitable for the location of new settlements or for intensifying the existing ones.

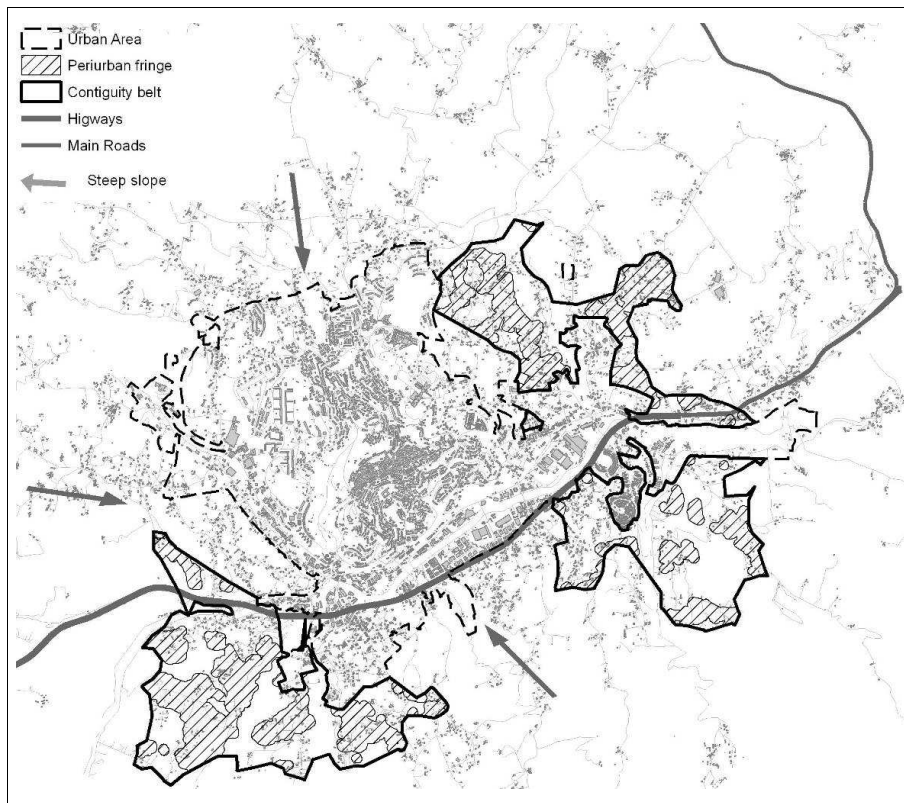


Figure 6. Periurban fringe after the land suitability procedure (oblique hatch).

After the theorization by Waldo Tobler (1970), the first law of geography is reported here: “Everything is related to everything else, but near things are more

related than distant things”. More experience exists of the use of spatial statistics in geographical analysis; for instance Kernel density has been applied for the location of epidemics (Gatrell *et al.* 1995) and studies on spreading of city services (Borruso and Schoier, 2004), while these techniques have not been used enough in the field of territorial planning. In this paper several kinds of spatial statistic functions have been applied for a deeper knowledge of territory and to give urban planners a better support for planning choices.

5. References

- Alberti M., Solera G., Tsetsi V., (1994), *La città sostenibile*, F. Angeli,
- Anselin, L., (1995), *Local Indicators of Spatial Association – LISA*. Geographical Analysis 27:93-115.
- Bailey T. C., Gatrell A. C. (1995), *Interactive spatial data analysis*, Prentice Hall
- Borruso G., Schoier G. (2004), *Density Analysis on Large Geographical Databases. Search for an Index of Centrality of Services at Urban Scale*, in Gavrilova M. L., Gervasi O., Kumar V., Laganà A., Mun Y. and Tan K. J. (eds.), Lecture Note in Computer Scienze Springer-Verlag, Berlin,
- Camagni, R., Capello R.; Nijkamp P., (1998), *Towards sustainable city policy: an economy-environment technology nexus*, Ecological Economics Volume: 24, Issue: 1, Elsevier
- Camagni, R., Gibelli M. C., Rigamonti P. (2002), *Urban mobility and urban form: the social and environmental costs of different patterns of urban expansion*, Ecological Economics Volume 40, Issue: 2, Elsevier
- Cavaillès J., Peeters D., Sekeris E., Thisse J.,(2004), *The periurban city: why to live between the suburbs and the countryside*, Regional Science and Urban Economics, Volume 34, Issue 6, Elsevier
- Gatrell A. C. , Bailey T. C., Diggle P. J., Rowlingson B. S., (1995), *Spatial point pattern analysis and its application in geographical epidemiology*, Transaction of institute of British Geographer
- Getis, A. and Ord, J. K., (1992), *The analysis of spatial association by use of distance statistics*, Geographical Analysis, 24, 189-206.
- Guerois M., Pumain D., (2001), *Urban sprawl in France (1950-2000)*, Franco Angeli
- Hidding M., Needham B., Wissershof, J., (2000), *Discourses of town and country*, Landscape and Urban Planning Volume: 48, Issue: 3-4, Elsevier
- Moran, P. (1948), *The interpretation of statistical maps*, Journal of the Royal Statistical Society, n.10
- O'Sullivan D., Unwin D., (2002), *Geographic Information Analysis*, John Wiley & Sons
- Van Den Berg L., Wintjes A. (2000), *New rural lifestyle estates in The Netherlands*, Landscape and Urban Planning, Volume: 48, Issue: 3-4 Elsevier

A New Information System for Tracing Geolocations of Bovine Cattle

Marcos Visoli*—Sonia Ternes*—Jean-Pierre Chanet**—François Pinet** —Andre Miralles*** —Stephan Bernard**—Gil Souza**

*Embrapa Agriculture Informatics, Campinas, Brazil

**Cemagref, Clermont-Ferrand, France

***Cemagref, Montpellier, France

ABSTRACT. The present paper describes the main idea of a new information system architecture dedicated to the animal geolocation acquisitions. It is applied to traceability of beefs in Brazil. In case of sanitary alerts, the system should be able to determine the animals which have been in contact with a diseased animal. This proposal is issued from studies undertaken in the project OTAG¹⁵ supported by the European Union. OTAG focuses on improving methods and geotechnologies for recording reliable and accurate data on beef production. OTAG develops an operational geo-decisional system to track and trace the mobility, provenance, and state of beef cattle. The presented method enables the acquisition of animal geolocations at a large scale. It minimizes the quantity of devices equipped on animals, and consequently the global economical and energetic costs of the system.

KEY WORDS: AGRICULTURE, BOVINE TRACEABILITY, GPS, UHF

¹⁵ Operational management and geo-decisional prototype to Track and trace Agricultural production - www.otag-project.org

1. Introduction

The main goal of this application paper is to sum up the first studies undertaken during a project supported by the European Community about geolocation and agricultural traceability.

The present paper introduces a first proposal of architecture for animal geolocation acquisitions at a large scale. It is applied to traceability of beefs in Brazil. This proposal is issued from the project OTAG (Operational management and geo-decisional prototype to Track and trace Agricultural production). The project is a specific support action of the 6th European Framework Programme. It concerns the bovine traceability in the context of Southern Cone Countries and EU policies. During the project, we proposed a new generation of methods to make the bovine traceability more reliable. The main idea of the solution is to use the last technologies of communication in order to memorize a precise history of animal geolocations. With the proposed system, deciders will have relatively exhaustive information about contacts between animals. In case of sanitary alerts (e.g. epidemic disease), they can deduce and delimitate very precisely the animals and the geographical areas (farms, pastures, etc.) having a potential risk. A goal of the solution presented in this paper is to minimize the economical and energetic costs of the geolocation acquisition.

In the last years Brazil has become the most important producer and exporter country of beefs in the world, with 2 millions tons in 2007. However, to maintain itself on this important economic position, Brazil needs constantly to improve its production systems as well to invest in aspects related to sanitary control. Brazil must fulfill the new requirements of markets such as the ones of European Community. The Brazilian beef production is developed all over the territory. It includes a population of 165 million of animals, and approximately 225 million hectares distributed in 2.20 million properties (IBGE, 2007).

The next section presents our analysis of existing traceability systems, and set the place of our proposal. Section 3 introduces the general architecture of our system and the main actors we consider; the possible use cases of the system has been studied for each actor. Section 4 describes our main idea for acquiring geolocations of animals. The presented method enables the acquisition of animal geolocations at a large scale. It minimizes the quantity of devices equipped on animals, and consequently the global economical and energetic costs of the system. Section 5 explains some models of databases that will be used to store animal geolocations.

2. Bovine traceability: existing systems and new needs

In Brazil there are two information systems related to beef cattle production. SISBOV (traceability service of beef cattle production chain) contains general information about the animal life inside the farm, from birth to death e.g. date of

birth, vaccinations (SISBOV, 2008). The farms who intend to export their production must use this system. To control the animal movement, Brazilian Ministry of Agriculture, Food and Livestock (MAPA) uses GTA system (Animal Traffic Guide). GTA is used to record places of origin, as well as farm entry/exit dates of animals. The information managed by SISBOV and GTA systems are complementary.

In Quebec, the Ministry of Agriculture has entrusted Agri-Traçabilité Quebec (ATQ) with developing and managing livestock identification and traceability system (AGROTRACABILITE, 2008). Quebec's livestock identification and traceability system allow knowing the main locations of each animal, its history, its transit and the animals which have been in contact with it. There are 3 types of information stored in the national ATQ database:

- the animal identification (thanks to an RFID ear tag),
- the site identification (an unique seven-digit number for each site, like farm, slaughterhouse and so on),
- the recording of each animal movement (and some additional events).

In Europe the regulation for animal identification, which is the responsibility of the Ministry of Agriculture of each individual EU country, is based on 4 principles:

- allocation of an unique national number for each animal,
- recording of the information regarding to the identity of every bovine into a cattle register held (recording made by producers for each livestock),
- communication of this information to a national computerised database,
- production of an identity document (cattle passport) to accompany every bovine travel.

These rules are the same in all the EU countries with some local adaptation. There are official procedures to make the equivalence of the data from different country.

In France, the Ministry of Agriculture delegates the practical running of the identification process to an official regional body (Etablissement départemental de l'élevage). The database for the registration of bovine in France is the BDNI (Base de Données Nationale d'Identification). It was created in 1999 by the Ministry. In January 2008, there was around 19,900,000 bovines registered and 246,000 farms. In France, there is also a project named GIEA which is focused on the data management of the farm (GIEA, 2008; DUFY et al., 2006; BRUN et al., 2005; POYET et al., 2003). A goal of this project is to propose a standard data format to facilitate information exchanges between agricultural software and French information systems. Once implemented in the main software applications used by farmers, this standard could be employed to encode data before their transmission

between farms and national information systems. An UML modelling of the GIEA data format is available (PINET et al., 2009 ; BOOCH et al., 1999).

Traditional systems are able to memorize the main bovine movements (e.g. from a farm to another), or to capture the geolocation of an animal when a transporter scans its RFID tag with a specific device (see ATQ systems). Now it should be possible to envisage a new generation of bovine traceability systems in which the different locations are automatically recorded several times per day for each animal. Memorizing the precise history of animal geolocations will improve considerably the traceability especially for countries having very large pasture areas (such as Brazil). With these new systems, deciders will have relatively exhaustive information about contacts between animals. In case of sanitary alerts, they can deduce and delimitate very precisely the animals and the geographical areas (farms, pastures, etc.) having a potential risk.

An objective of OTAG is to propose a technological solution and a global architecture for a new generation of animal traceability systems. For measuring the animal geolocations in the pastures, the project proposed the use of electronic devices embedded in necklaces carried out by beefs. Then these data will be centralized in databases at different levels. These records will constitute a valuable information source for providing an efficient traceability of animals, and as a consequence, a better sanitary control and a reliable alimentary security.

3. OTAG solution: architecture layers and identified actors

The OTAG architecture is organized in four layers as schematized in Figure 1. The layer 1 concerns the acquisition of georeferenced data in the agricultural pastures; the data are captured thanks to embedded electronic devices carried out by the animals. The layer 2 is responsible for storing collected data at the farm level in using local information systems and databases. In the layer 3 the information of all the farms supporting our system are centralised in a same transactional database. The layer 4 is dedicated to the geo-decisional analysis; its main goal is to help public authorities make decisions (for instance in case of sanitary alerts). This last layer could be supported by different geo-decisional tools and a data warehouse that will integrate the database presented in layer 3 with several external sources of data e.g. information about ground, pastures, climate, etc.

We propose bellow the use cases of the possible protagonists involved in the system.

Government. Data about animal geolocations can be analysed by public authorities in order to make decisions in different situations when a sanitary alert occurs.

Consumers. The consumers are the domestic market of bovine meat. Our system could give information for those actors about production system practices and about the animal life and its origin. Those data contribute for increasing product quality.

Researchers. The researchers could get many information from the system, to compare the results of different production systems used in farms, to analyse data from different sources in order to study the real impact of grounds, pastures and climate in the bovine production systems and to analyse in a more realistic way the relation between animal movement and illness occurrence.

Farmers. Farmers are the most important actors in the system. The quality of information depends directly on the benefits that they could get by using the system.

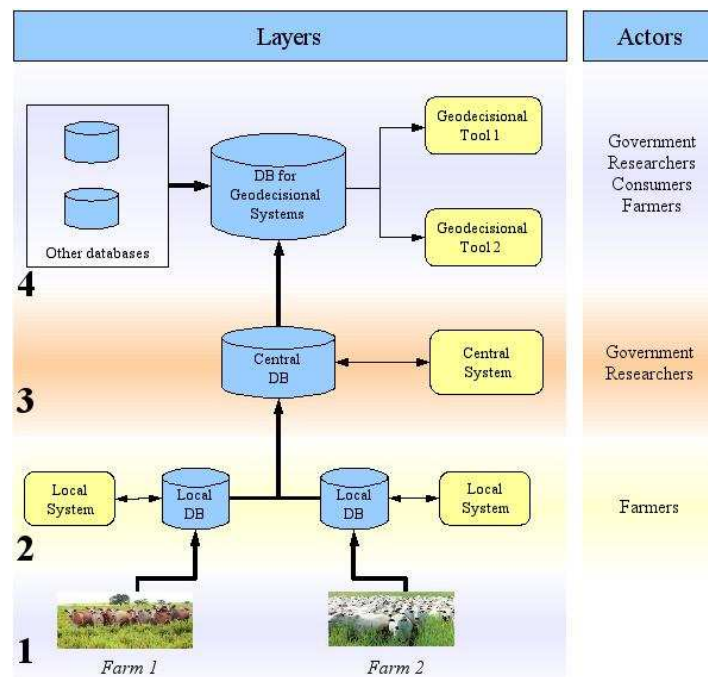


Figure 1. Layers of the traceability architecture

4. New principles of data acquisition

An innovative mechanism has been proposed to acquire geolocations of animals. In the proposed system, geolocations of animals will be measured periodically, by the use of specific electronic necklaces on the animals, and by considering the concept of “master” and “slave” animals.

In a herd, several animals follow the same leader animal. In fact, there are groups of animals and all the animals inside the group follow the same leader. In the proposed system the leader animal will have a specific “master” necklace and will

capture periodically the identification of the “slave” necklaces of all animals close to it.

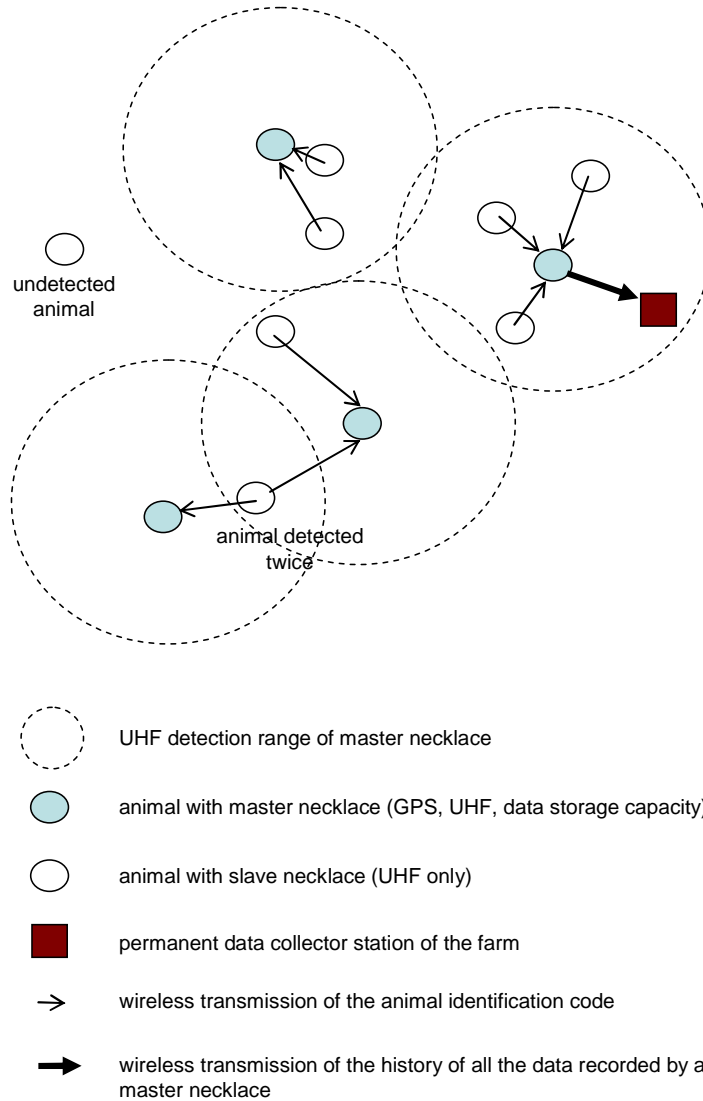


Figure 2. *New principles for the acquisition of animal geolocations*

The animals that are used to be leader must be beforehand identified in the herd by the farmer and they will be fitted out with a specific “master” necklace. This type of electronic necklace includes a GPS module, a data storage capacity, and a UHF (wireless) communication module. The other animals are equipped with a “slave”

necklace that includes only a UHF system that is used to signal their presence to the master animal. The master animals are able to detect the presence of neighbours in using their UHF modules. The test of detection of presence is initiated every hour by the master necklace; indeed, every hour, it sends a wake up call to the slave necklaces close to it. Once they received this signal, the slave necklaces send their identification code the master necklace. At that moment, the master necklace records all the identification codes it received, as well as its GPS position and the current time. This method allow recording inside master necklaces, the rough geolocations of the slaves necklace. Only the position of the master necklace is acquired by the GPS module, but one can consider that the real positions of the slave necklaces are not very different from the one of the master. All the data recorded by master necklaces are transferred automatically to a permanent data collector station when the master animals pass it. With this new method, the majority of animals have only a UHF module. This technique enables the acquisition of animal geolocations, and minimizes the quantity of devices equipped on animals, and consequently the global economical and energetic costs of the system (including the maintenance costs). Figure 2 sums up the proposed architecture. As illustrated in the Figure, while some isolated animals can remain undetected, some others can be detected twice. The collected geolocation data will be centralized in a local database at the farm level, together with the traceability of different other information concerning the health of animals (weight, vaccination, body temperature, etc.).

5. Design of local systems

The model that describes data of local systems (layer 2 in Figure 1) has been formalized with UML. This type of representation is the better means to describe precisely data handled by local systems. The model has been built iteratively thanks to meetings and discussions between the different researchers involved in the project. Local systems are based on concepts like farm, farmer, cattle area, paddock, animal, sanitary event and geolocations. These concepts come from practices of Brazilian beef cattle production systems.

The diagrams of Figure 3 and 4 are subparts of the proposed conceptual model of the local database; they sum up the main data. The farm is an agricultural establishment that belongs to one owner. Each cattle area is under the responsibility of a farmer, that can be the owner or a leaseholder. Each cattle area corresponds to a geographical zone with one specific production system (extensive, semi-intensive and intensive), that can change periodically. This area is divided into paddocks; the paddock is associated with a feeding system, that can be modified throughout the time. A group of animals can be moved from a paddock to another one. The animals periodically receive vaccines. Their weight and their body temperatures are periodically observed. Moreover, different fundamental events are recorded: birth, death, entrance and exit of the farm. The positioning of the animals is measured periodically thanks to necklaces.

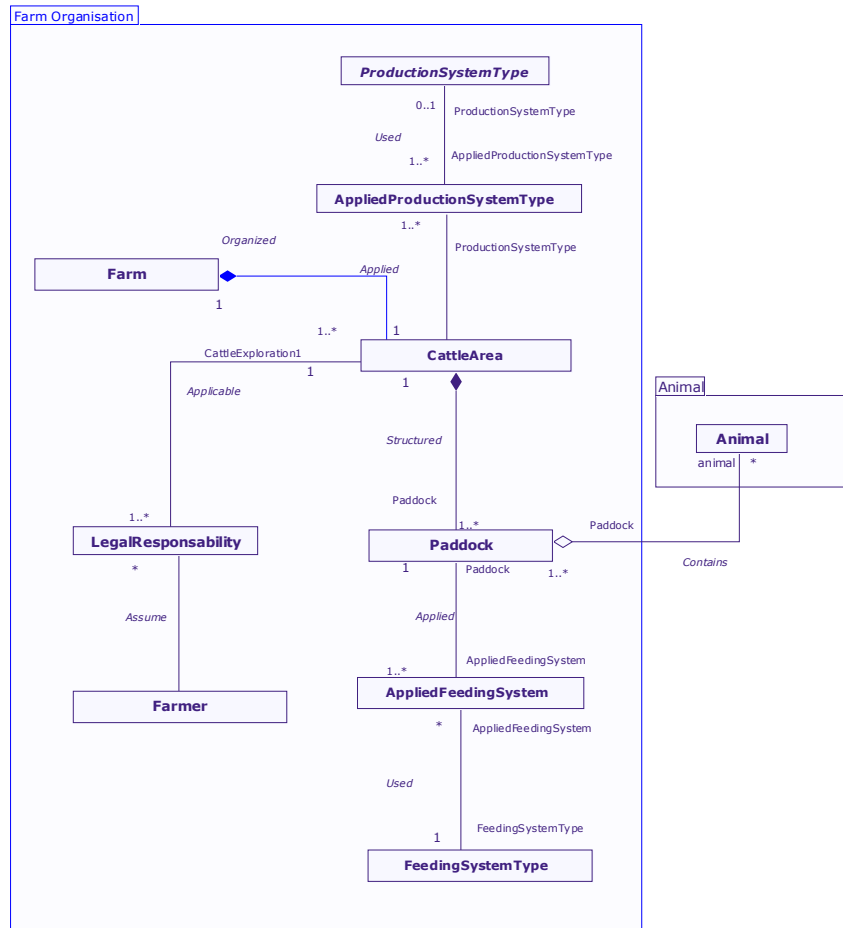


Figure 3. Main concepts of the local databases (Part 1)

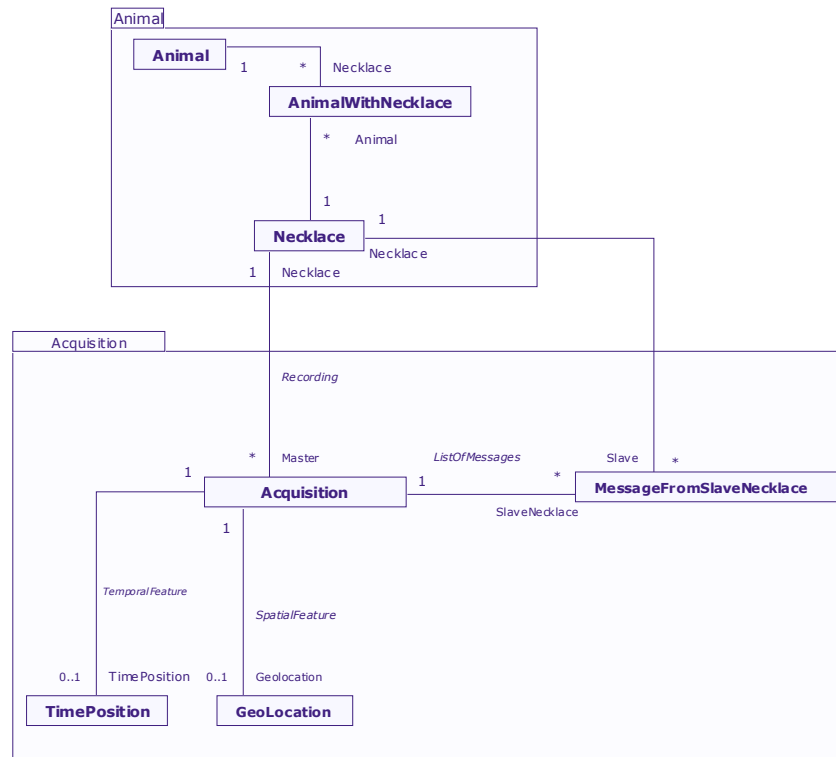


Figure 4. Main concepts of the local databases (Part 2)

6. Conclusion

OTAG focuses on an important target that involves a large number of actors around an proprietary subject for the human health. The development of the operational prototype is a very important requirement to move towards an efficient system for the bovine traceability. The system should be useful by governmental institutions in real situations.

Currently, the main architecture of the system has been proposed. The first version of the model of the database of local systems has been built, and an hardware prototype of necklace has been setting up (OTAG REPORT, 2008). A specific code has also been implemented to simulate the moving of groups of animals; it could be used to estimate the evolution of the number of non-detected animals depending on the number of animal fitted out with a master necklace. The

next steps concern mainly the development of real experiments with several necklaces, and the specification of geo-decisional analyses.

At the level of layers 3-4 (Figure 1), in order to get an acceptable response time for answering user queries, it will be necessary to organize the great amount of collected data. The possibility of integration with external databases is studied. This can contribute for a better analysis aiming the decision making about the farm production. For instance, the integration of OTAG prototype with databases about ground, pastures and climate can increase the quality of information concerning to beef cattle production.

7. Bibliography

- AGROTRACABILITE, 2008. Site Web Agi-traçabilité Québec, <<http://www.agri-tracabilite.qc.ca>>
- Booch G., Rumbaugh J., Jacobson I., 1999. The Unified Modeling Language User Guide, Addison-Wesley.
- BRUN T., POYET P., BOOP M., VIGIER F., 2005. Towards an agricultural ontology in France: Contributions of the farm Information Management Project (GIEA), EFITA/WCCA 2005 Joint Conference, Vila Real, Portugal, p. 1296-1302.
- DUFY, L., ABT, V., POYET, P., 2006. GIEA : gestion des informations de l'exploitation agricole. Un projet au service de l'interopérabilité sémantique de la profession agricole. Ingénieries - Eau, Agriculture, Territoires n° 48, p. 27 – 36
- GIEA, 2008. Site Web GIEA, <<http://www.projetgiea.fr>>
- IBGE – Instituto Brasileiro de Geografia e Estatística, 2007. <www.ibge.gov.br>
- OTAG REPORT, 2008. Traceability in beef production and crisis management in bovine sector: state of art. OTAG Project Report (WP2).
- PINET F., ROUSSEY C., BRUN T. , VIGIER F., 2009. The Use of UML as a Tool for the Formalisation of Standards and the Design of Ontologies in Agriculture. Chapter in: Advances in Modeling Agricultural Systems, Springer.
- POYET, P., BRUN, T., 2003. GIEA : gestion des informations de l'exploitation agricole - Vers des concepts et un langage communs pour les partenaires économiques et institutionnels de la profession agricole. Ingénieries - Eau, Agriculture, Territoires, n° spécial Technologies pour les agrosystèmes durables, p. 167 – 175
- SISBOV, 2008, SISBOV web site, <<http://www.localizaanimal.com.br/>>

Apprentissage incrémental de connaissances à partir de données d'un simulateur

Marie-Odile Cordier, Thomas Guyet, Christine Largouët, Véronique Masson, Henri-Maxime Suchier

*Université Rennes-1/IRISA
Campus de Beaulieu 35042 Rennes Cedex
cordier@irisa.fr*

*AGROCAMPUS OUEST
65 rue de St Briec, CS 84215, 35042 Rennes Cedex
christine.largouet@agrocampus-ouest.fr*

RÉSUMÉ. Cet article présente un travail en cours sur une méthode incrémentale et interactive d'apprentissage de règles de classification à partir de données de simulation. Nous cherchons à améliorer une méthode existante d'apprentissage de règles de classification à partir des données d'un simulateur. Ce simulateur, développé dans le cadre du projet SACADEAU, porte sur l'écoulement des eaux et des polluants à l'échelle d'un bassin versant. Nous présentons les principes d'une méthode incrémentale et interactive en proposant des solutions préliminaires de stratégies pour les étapes clés de la méthode. Ces solutions préliminaires sont évaluées avant d'aborder les perspectives d'amélioration.

ABSTRACT. This article presents a work in progress on an incremental and interactive learning method of classification rules from simulation data. Using this technique, we expect to improve the results of an existing approach, developed in the SACADEAU project and which learns classification rules from water and pollutants flows throughout a catchment area. We present the principles of an incremental and interactive method and we propose preliminary strategies for the key stages. The method is evaluated with these preliminary solutions and we give the prospects to explore for future improvements.

MOTS-CLÉS: Approche incrémentale, apprentissage automatique, règles de classification

KEYWORDS: Incremental approach, machine learning, classification rules

1. Introduction

L'objectif du projet SACADEAU (Système d'Acquisition de Connaissances pour l'Aide à la Décision pour la qualité de l'EAU) est de construire un outil d'aide à la décision destiné aux gestionnaires d'un bassin versant afin de les aider et préserver la qualité de l'eau à l'exutoire. Deux fonctionnalités sont attendues de cet outil : d'une part, permettre de tester par simulation différents scénarios afin d'évaluer l'impact des stratégies de désherbage, des configurations spatiales des parcelles de maïs traitées, des aménagements de l'espace rural (haies, bandes enherbées) ; d'autre part, analyser l'ensemble des résultats de simulation obtenus afin de déterminer les variables explicatives, de mieux comprendre leur rôle et leur importance relative dans la contamination des eaux, et éventuellement de pouvoir recommander des actions améliorant les situations courantes (Cordier *et al.*, 2005).

Dans cet article, nous nous intéressons au second point et poursuivons un travail effectué dans le cadre de la thèse de R. Trépos (Trépos, 2008) sur l'aide que peut apporter les méthodes d'apprentissage automatique de l'intelligence artificielle pour l'analyse des résultats de simulation obtenus grâce au modèle. Tirant parti de la spatialisation du bassin versant, et de la notion d'arbre d'exutoire, R. Trépos construit un ensemble d'exemples constituant la base d'apprentissage. L'approche CN2 (Clark *et al.*, 1989), qui ne sera pas détaillée dans cet article, permet d'obtenir des règles de classification de type attribut-valeur généralisant ces exemples. Les règles de classification permettent de lier des facteurs agronomiques au transfert de pesticides. On cherche à extraire des règles qui discriminent les situations de transfert important de celles de transfert faible. C'est à partir de ces règles de classification que des règles de décision pourront être construites. Seul l'aspect portant sur les règles de classification est abordé dans cet article.

Sans remettre en cause l'intérêt de ce premier travail, nous l'avons poursuivi avec l'idée d'utiliser ces mêmes techniques d'apprentissage de manière incrémentale et interactive. L'utilisateur, pour nous le gestionnaire du bassin versant, soumet une requête au système qui, suite au processus d'apprentissage fournit une certaine généralisation des exemples. Les réponses apportées peuvent conduire l'utilisateur à souhaiter des précisions, compléter ou élaguer des règles. Il entre alors dans une boucle d'interaction incluant le simulateur et le processus d'apprentissage. Nous avons donc démarré un travail sur l'apprentissage incrémental qui fait l'objet de cet article.

Partant d'un ensemble initial restreint d'exemples, l'apprentissage de type CN2 est effectué et fournit un ensemble de règles de classification. Leur qualité est évaluée en utilisant des critères objectifs (spécificité, précision ...) et des critères subjectifs (reflétant leur adéquation à la requête de l'utilisateur). Les meilleures d'entre elles sont conservées dans une base de règles. Ensuite, une analyse de la base de règles, obtenue à cette première itération, conduit à modifier les paramètres de simulation et une nouvelle itération est lancée. L'ensemble de règles est ainsi construit de manière incrémentale, en tenant compte à chaque itération des résultats obtenus et de la de-

mande de l'utilisateur. Ce cycle est poursuivi jusqu'à obtention d'une base de règles ayant atteint une qualité suffisante.

Après un rappel du contexte par la description succincte du projet SACADEAU en section 2, nous présentons les principes et l'architecture du prototype que nous avons développé en section 3.1. Les premiers résultats obtenus sont donnés en section 4 avant de conclure en ouvrant les perspectives de ce travail en cours de développement.

2. Projet SACADEAU

Nous rappelons dans cette section quelques éléments permettant de comprendre le contexte du travail présenté dans cet article.

Le projet SACADEAU (Cordier *et al.*, 2006) s'est focalisé sur une application portant sur le désherbage du maïs et son impact sur la contamination des eaux superficielles dans des contextes de nappe superficielle liée à la topographie, et de systèmes agricoles liés à l'élevage où la part du maïs dans l'assolement est importante. Le site d'étude est celui du bassin versant du Frémur (Morbihan) qui est un des bassins versants du programme Bretagne Eau Pure.

L'architecture générale est donnée dans la figure 1.

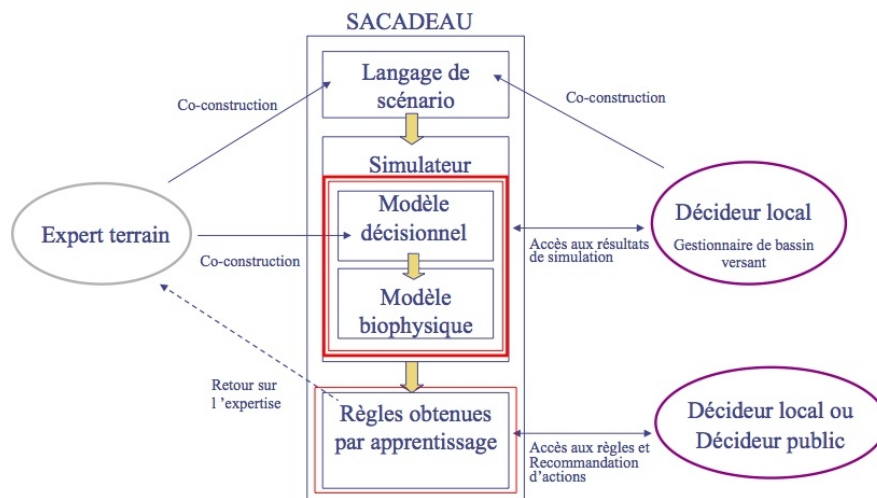


Figure 1. Architecture générale du projet SACADEAU.

Le modèle (*cf.* figure 2) évalue la contamination des eaux à l'exutoire de bassins versants de taille moyenne (quelques kilomètres carrés), au pas de temps journalier, sur la période proche des applications. Il permet, à partir de données climatiques, de

la configuration spatiale des cultures et des applications, de stratégies et de modalités de désherbage des cultures (appelés également *itinéraires techniques*, et noté ITK), d'obtenir par simulation les niveaux de contamination de l'eau à l'exutoire d'un bassin versant, jour après jour, sur les quelques mois suivant les applications. Ce modèle comprend quatre sous-modèles : un modèle biophysique de transfert des herbicides, un modèle climatique générant des données journalières, un modèle décisionnel générant les données relatives aux interventions agricoles, enfin un modèle spatial générant les données relatives à la topologie du bassin versant.

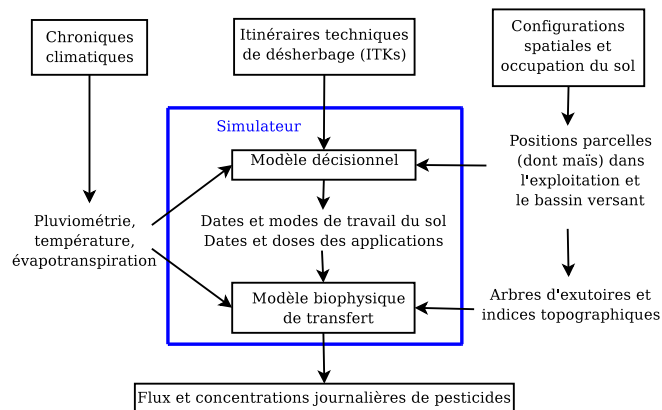


Figure 2. Le modèle SACADEAU est le couplage d'un modèle décisionnel simulant les pratiques agricoles et d'un modèle biophysique des transferts de l'eau et des herbicides.

La modélisation s'appuie une représentation spatiale du bassin versant (Aurousseau *et al.*, 2009) utilisant la notion d'arbre d'exutoire. Un arbre d'exutoire représente les chemins de l'eau connectant un ensemble de sous-parcelles contribuant à un exutoire. Chaque arbre d'exutoire est indépendant d'un point de vue du ruissellement puisque aucun flux n'est échangé entre deux exutoires. Les résultats de simulation du modèle décrivent les flux d'eau et de pesticides pour chaque arbre d'exutoire du bassin versant. À chaque arbre d'exutoire est associé un certain nombre d'attributs décrivant les caractéristiques des parcelles contributives ainsi qu'un attribut de sortie qui est le taux de transfert. Celui-ci exprime la résistance du paysage au transfert d'eau et de charges polluantes.

L'apprentissage s'effectue à partir d'une base d'apprentissage, collectant les résultats associés aux arbres d'exutoires des différentes simulations. Deux types d'apprentissage ont été expérimentés. L'approche que nous retenons consiste à définir un ensemble d'attributs synthétisant l'information contenue dans les arbres d'exutoires et utilise ensuite l'algorithme CN2 pour produire un ensemble de règles de classification attribut-valeur.

Par exemple, la règle suivante prédit un transfert important pour un arbre d'exutoires possédant au moins un exutoire de parcelle maïs de taille importante (au moins 1,54 ha) et l'utilisation de matières actives à risque fort de transfert (au moins 50% de la quantité totale appliquée).

```
SI agr_surf_max_maïs > 1,54 ha
ET agr_rapport_risque_fort > 50 %
ALORS classe = transfert_important [support=125,confiance=1]
```

3. Présentation du système

3.1. Principes

Dans ce travail, l'apprentissage itératif est utilisé pour construire progressivement un ensemble de règles de classification. L'objectif consiste à proposer au final un ensemble de règles avec une qualité supérieure à celle qui aurait été obtenue par un apprentissage unique. Le principe repose sur l'amélioration de la qualité de cet ensemble de règles, en jouant à plusieurs reprises les étapes simulation-apprentissage.

Le système mis au point dans le projet SACADEAU repose sur deux étapes (*cf.* figure 3) : la mise en œuvre du simulateur qui fournit un jeu d'exemples et l'apprentissage proprement dit qui produit un ensemble de règles de classification. On considère par la suite que ces deux étapes constituent une *itération*. On appelle *configuration* les variables en entrée du simulateur qui affectent un des 62 itinéraires techniques (ITK) à chaque exploitation agricole du bassin versant (les variables climatiques et les configurations spatiales restent inchangées, *cf.* figure 2). Ces paramètres sont définis manuellement selon le type de simulation recherchée. Le simulateur fournit en sortie des exemples représentant des arbres d'exutoires répartis en deux classes : ceux ayant un transfert important de pesticide et ceux ayant un transfert faible. La technique d'apprentissage CN2 induit un ensemble de règles de classification. Les résultats peuvent être enfin évalués à partir des exemples tests de la base d'apprentissage.

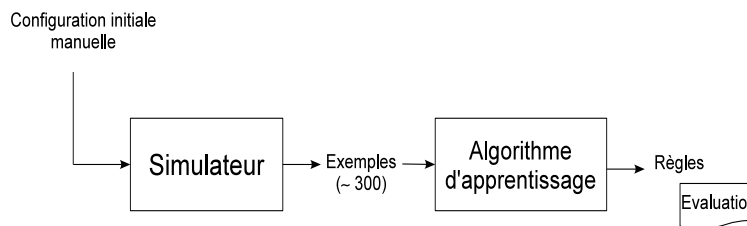


Figure 3. Apprentissage en une "itération" (projet SACADEAU)

Dans le cadre de l'approche décrite dans ce papier et présentée figure 4, l'*itération* se répète jusqu'à ce que l'ensemble de règles atteigne une qualité satisfaisante pour

l'utilisateur. À la suite de l'étape d'apprentissage, les règles obtenues sont évaluées de sorte à sélectionner automatiquement celles qu'il faudrait ajouter à la théorie globale. Toutes les règles ainsi que leurs critères d'évaluation sont présentés à l'utilisateur. Celles qui ont été sélectionnées lui sont suggérées, mais il qui dispose de toute la latitude pour intégrer une ou plusieurs règles à la base de règles, appelée *théorie globale*. Une fois cette étape de sélection des règles achevée, il s'agit de préparer l'itération suivante en déterminant la nouvelle configuration du simulateur. Une fois la nouvelle configuration définie, l'itération suivante peut alors démarrer et les itérations s'enchaînent jusqu'à satisfaction de l'utilisateur vis-à-vis de la théorie globale obtenue.

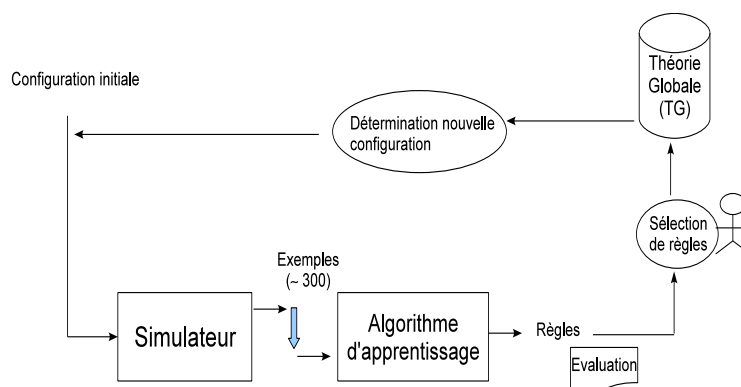


Figure 4. Apprentissage selon le principe incrémental et interactif

Détaillons plus précisément la méthode permettant d'établir une nouvelle configuration. Les 62 ITK sont regroupés selon 6 types d'ITK. Une nouvelle configuration est générée aléatoirement à partir d'une distribution de la probabilité qu'un exemple¹ utilise un ITK de l'un des 6 types. Initialement, on utilise une distribution équiprobable d'affectation d'un ITK à un exemple.

3.2. Choix de stratégies

Les choix de stratégie pour les deux étapes clés : 1) la détermination d'une nouvelle configuration, 2) la sélection automatique de règles à ajouter, sont détaillées ci-dessous. D'autres stratégies reposant sur des critères différents peuvent être envisagées et sont décrites section 5.

Pour la détermination d'une nouvelle configuration, nous utilisons un critère de couverture. À la suite d'une itération, nous nous intéressons à l'espace des ITKs non

1. En fait, un exemple (arbre d'exutoire) ne peut être directement associé à un ITK. On fait ici l'hypothèse que statistiquement la distribution des arbres d'exutoires *représentatifs* d'un type d'ITK sera bien proportionnel à la distribution de probabilité des types d'ITK.

couverts par la théorie globale, avec pour objectif de simuler des exemples correspondant à cet espace dans les itérations suivantes. Notre stratégie consiste tout d'abord à repérer la nouvelle zone de l'espace de recherche à explorer par un point de référence. Ce point de référence sera ensuite utilisé pour adapter la distribution de probabilités. Pour construire le point de référence, un clustering (*K-Means* avec 6 classes) est appliqué sur les exemples non couverts par la théorie globale. Des attributs communs, représentant la pression d'herbicides, permettent de comparer les ITKs et les exemples (les arbres d'exutoires) selon un même espace à deux dimensions. Le centre du plus cluster comportant le plus d'exemples mals classés sert de point de référence pour l'espace à explorer. Le type d'ITK le plus proche du point de référence est privilégié lors de l'itération suivante. La détermination d'une distribution de probabilités des types d'ITK consiste à adapter les probabilités p_i de chaque type d'ITK $i \in \mathbb{N}_5$ (de centre C_i), de la manière suivante :

$$p_i^{new} = \frac{p_i^{old}}{\|P_r - C_i\|_2}$$

où $\|P_r - C_i\|_2$ est la distance euclidienne entre le point de référence et le centre du type d'ITK i . La nouvelle distribution est ensuite normalisée.

Pour la suggestion de règles à ajouter, notre stratégie de sélection de règles utilise la confiance² d'une règle (sur la base d'apprentissage). Toutes les règles générées à une itération qui ont une confiance supérieure à 80% sont ajoutées à la théorie globale.

4. Implémentation et résultats

Nous avons implanté un outil permettant à un utilisateur de construire incrémentalement et interactivement un ensemble de règles. Il utilise des réalisations existantes : le simulateur (Tortrat, 2006) et l'algorithme d'apprentissage CN2 (implémentation de Clark (Clark *et al.*, 1989)).

La figure 5 illustre l'interface graphique de l'outil. On se trouve ici à la fin du troisième cycle pendant lequel 3 nouvelles règles (*cf.* figure 5, boîte de dialogue) ont été trouvées. Les règles surlignées sont celles qui sont suggérées automatiquement pour l'ajout à la théorie globale (l'utilisateur peut ou non suivre ces suggestions). La théorie globale comporte déjà 4 règles (*cf.* figure 5, partie supérieure gauche de la fenêtre en arrière plan).

La partie droite de l'interface sert à la visualisation des exemples d'apprentissage dans un plan défini par 2 attributs. D'autres outils de visualisation permettent de suivre l'évolution de la qualité de la théorie globale en construction.

Dans l'étape d'évaluation, nous cherchons à évaluer la pertinence des stratégies choisies. Pour cela, on a réalisé des apprentissages incrémentaux en s'arrêtant au bout

2. La confiance d'une règle est calculée comme le nombre d'exemples de la classe d'intérêt de la règle parmi tous les exemples couverts par la règles.

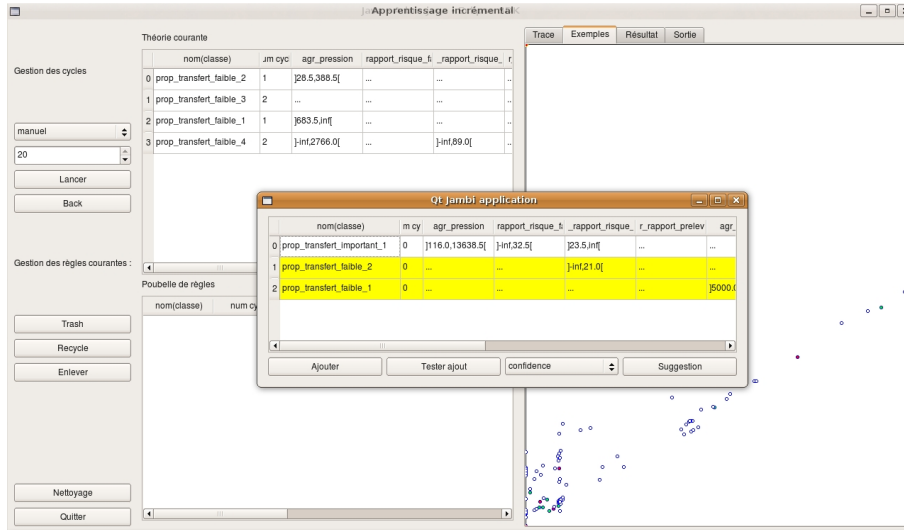


Figure 5. Interface (cf. texte pour les détails).

de 20 itérations³. Nous n'évaluons pas ici l'intérêt de l'aspect interactif de la méthode, mais uniquement l'aspect incrémental. En pratique, les règles choisies par la stratégie de sélection de règles sont celles qui sont ajoutées à la théorie globale.

À chaque étape, on calcule la précision de la théorie globale sur une base de validation constituée d'un grand nombre d'exemples (1300 dans notre application). La précision est calculée comme le nombre d'exemples bien classés par la théorie globale sur le nombre total d'exemples. On construit ainsi des courbes d'évolution de la précision en fonction des itérations successives.

La figure 6 illustre un exemple d'évolution de la précision en fonction des itérations successives. Si on reproduit l'expérimentation plusieurs fois, on obtient des courbes assez similaires. Mais toutes ne sont pas toujours croissantes. En moyenne, la précision maximale atteinte est de $0.83(\pm 0.04)$ et la précision moyenne est de $0.81(\pm 0.03)$. Pour comparaison, les résultats de l'apprentissage non incrémental ont permis d'identifier 33 règles avec un taux de confiance de 88%.

À la fin des 20 itérations, le système a construit une théorie globale qui comporte en moyenne 46.3 règles. Le nombre de règles ajoutées à chaque itération est stable (2.32 ± 0.57). Une grande majorité des règles (>90%) ont pour conclusion le transfert faible. La variabilité des arbres d'exutoires générés ayant ce comportement est donc certainement plus important que celle des arbres ayant un transfert important. La prémise d'une règle comporte en moyenne $3.56(\pm 0.94)$ couples attributs-valeurs.

3. L'ensemble des 20 itérations est réalisé en environ 1min sur un Intel Centrino Duo.

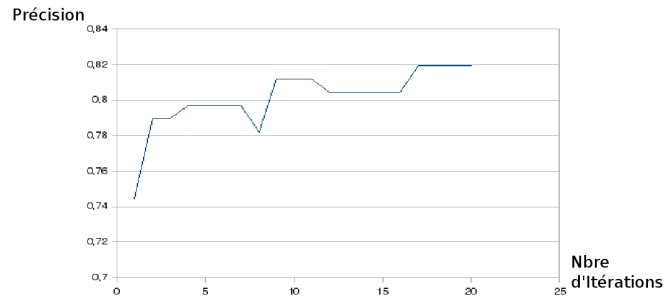


Figure 6. Exemple de courbe d'évolution de la précision en fonction des itérations successives.

5. Conclusion et perspectives

Dans ce travail exploratoire nous avons cherché à expérimenter une méthode d'apprentissage incrémentale et interactive afin d'améliorer un ensemble de règles de classification. Si ce type de techniques commence à apparaître dans la communauté de l'apprentissage automatique, c'est cependant, à notre connaissance, la première fois qu'un simulateur est couplé à un algorithme de généralisation afin de fournir de nouveaux exemples. Les résultats obtenus dans SACADEAU sont supérieurs à ceux issus de plusieurs itérations. Et nous travaillons actuellement sur des critères utilisés par les stratégies qui permettront d'améliorer ces résultats :

- les critères d'évaluation des règles pour la stratégie de sélection,
- les critères d'évaluation des zones de recherche à privilégier pour les stratégies de détermination de la configuration suivante.

Dans le cadre d'une procédure interactive telle que présentée ici, l'évaluation des règles au sein d'une théorie est un élément fondamental puisqu'il s'agit de fournir le maximum d'information à l'utilisateur pour l'aider à comprendre et à sélectionner les règles. Jusqu'à présent, seuls des critères statistiques classiques (confiance) ne prenant pas en compte le reste de la théorie sont utilisés, cependant d'autres critères de qualité pourraient être envisagés. En conséquence, le système trouve toujours de nouvelles règles à ajouter, mais celles-ci peuvent être partiellement redondantes et leur apport à la théorie globale est souvent mineur (précision inchangée). Dans le domaine de l'extraction des connaissances, ces critères sont classés en critères objectifs, *i.e.* guidés par les données, et critères subjectifs, *i.e.* guidés par l'expert du domaine. Concernant les critères objectifs (Lenca *et al.*, 2008) permettant d'évaluer les règles en vue d'une sélection : l'intérêt (qui s'appuie sur le nombre de contre-exemples), la redondance ou la similarité (par rapport au contenu actuel de la théorie globale), la forme des règles, la généralité ou la rareté seraient des critères intéressants à prendre en considération. Les mesures subjectives dépendent du point de vue de l'utilisateur et sont généralement

de deux types (Al-Hegami, 2004) : la surprise et l'actionnabilité. Vis-à-vis du critère de surprise, une règle sera intéressante si elle surprend l'utilisateur par rapport à ses propres connaissances du domaine. Vis-à-vis du critère d'actionnabilité, une règle sera intéressante si elle aide l'utilisateur à prendre une décision dans son travail. Bien que majeur dans un contexte d'aide à la décision tel que SACADEAU, l'actionnabilité reste un critère difficile à définir de façon formelle.

Le deuxième point clé à explorer est la stratégie de détermination d'une nouvelle configuration pour l'itération suivante. La difficulté repose dans la construction d'un lien entre les résultats en sortie de l'algorithme d'apprentissage et les variables en entrée du simulateur. Jusqu'à présent, nous avons utilisé un critère qui indique, parmi les 6 zones définies par les types d'ITK, la zone de l'espace de variables d'entrées qu'il faut explorer aux itérations suivantes. Des critères pouvant localiser plus finement l'espace des entrées à explorer sont donc à étudier.

6. Bibliographie

- Al-Hegami A., « Subjective measures and their role in data mining process », *Proceedings of the International Conference on Cognitive Systems (ICCS)*, 2004.
- Aurousseau P., Squividant H., Trépos R., Tortrat F., Gascuel-Oudou C., Cordier M.-O., « A plot drainage network as a conceptual tool for the spatialisation of surface flow pathways for agricultural catchments », *Computer & Geosciences*, vol. 35, p. 276-288, 2009.
- Clark P., Niblett T., « The CN2 induction algorithm », *Machine Learning*, vol. 4, n° 4, p. 261-283, 1989.
- Cordier M.-O., Garcia F., Gascuel C., Masson V., Salmon-Monviola J., Tortrat F., Trépos R., « A machine learning approach for evaluating the impact of land use and management practices on streamwater pollution by pesticides », *Proceedings of the International Congress on Modelling and Simulation MODSIM'05*, 2005.
- Cordier M.-O., Tortrat F., Trépos R., Aurousseau P., Chanomordic B., Falchier M., Gascuel-Oudou C., Garcia F., Heddadj D., Lebouille L., Masson V., *Modélisation du transfert des herbicides dans un bassin versant en vue de la construction d'un outil d'aide à la décision pour la maîtrise de la qualité des eaux*, Update Sciences & Technologies, Éditions Inra, p. 219-225, 2006.
- Lenca P., Meyer P., Vaillant B., Lallich S., « On selecting interestingness measures for association rules : User oriented description and multiple criteria decision aid », *European Journal of Operational Research*, vol. 184, p. 610-626, 2008.
- Tortrat F., *Modélisation orientée décision des processus de transfert par ruissellement et subsurface des herbicides dans les bassins versants agricoles*, PhD thesis, ENSA de Rennes, 2006.
- Trépos R., *Apprentissage symbolique á partir de données issues de simulation pour l'aide á la décision. Gestion d'un bassin versant pour une meilleure qualité de l'eau*, PhD thesis, Université Rennes-1, 2008.

Une démarche fondée sur les treillis de Galois pour l'aide à la qualification de l'état des milieux aquatiques

Agnès Braud* — **Corinne Grac**** — **Silvia Pristavu*** — **Emilie Dor****
— **Florence Le Ber**,****

* *LSIT UMR 7005, Bd Sébastien Brant, BP 10413, 67412 Illkirch cedex FRANCE*
agnes.braud@unistra.fr

** *LHyGeS UMR 7517 - ENGEES, 1 quai Koch, 67000 Strasbourg FRANCE*
{corinne.grac, florence.leber}@engees.u-strasbg.fr

*** *LORIA UMR 7503, BP 35, 54506 Vandœuvre-lès-Nancy cedex FRANCE*

RÉSUMÉ. L'évaluation de la qualité des cours d'eau est un problème majeur et complexe. Différents paramètres apportent des informations sur l'état de l'eau parmi lesquels des indices biologiques. Cinq indices ont été normalisés en France, mais leur interprétation est difficile, et plus particulièrement la corrélation de ces indicateurs. Le travail présenté dans cet article a pour objectif de déterminer des profils de stations (portions de cours d'eau) caractérisés par des valeurs pour ces indices. Nous proposons d'utiliser les treillis de Galois, comme outil d'extraction de connaissances permettant d'accompagner les biologistes dans l'explicitation de leur expertise et de les aider à mettre en évidence l'ensemble des interactions entre les compartiments biologiques. La finalité est d'intégrer ces profils dans un classifieur capable de qualifier une nouvelle station en fonction de ses valeurs pour les indices. La démarche complète est expliquée et les premiers résultats obtenus par ce processus sont détaillés. Ils montrent l'intérêt de cette approche pour l'identification de profils de stations pertinents par les biologistes.

ABSTRACT. The evaluation of water quality is a major concern and a complex task. Several parameters bring information about this quality among which biological indices. Five indices have been normalized in France, but their interpretation is difficult, and more precisely their correlation. The work presented in this article aims at determining profiles for water stations (part of water bodies) characterized by values for those indices. We propose to use Galois lattices as a knowledge extraction tool to help biologists in making explicit their expertise and the interactions between biological compartments. The final goal is to integrate those profiles in a classifier allowing to qualify a new station according to its values for the indices. The whole

process is explained and some results obtained on our data are detailed. They show the benefits for the biologists of using this approach to identify relevant station profiles.

MOTS-CLÉS : Treillis de Galois, concepts formels, règles d'implication, état biologique des masses d'eau, indices biologiques

KEYWORDS: Galois lattices, formal concept, implication rules, biological water quality, biological indices

1. Introduction

L'évaluation de la qualité des cours d'eau est un problème majeur et complexe. La mise en place de mesures de contrôle et de sauvegarde de l'environnement, demandée notamment par la directive-cadre européenne sur l'eau (DCE), nécessite la formalisation de cette évaluation.

Différents paramètres apportent des informations sur l'état des cours d'eau : des paramètres biologiques (indiquant l'état des flore et faune aquatiques), physico-chimiques, physiques (par exemple l'état des berges). En France, la physico-chimie est mesurée très régulièrement sur ces réseaux. Elle peut permettre de détecter une pollution, mais reflète un état ponctuel et se révèle donc insuffisante comme outil d'évaluation. Au contraire, les populations présentes dans le milieu, information capturée par les paramètres biologiques, ne rendent pas compte immédiatement d'une pollution mais sont ensuite affectées durablement. Cinq indices biologiques ont été définis en France. Ils portent chacun sur un groupe taxonomique particulier et se basent sur les taxons de ce groupe présents dans l'eau. Ces groupes sont les macrophytes et les diatomées (flore), les invertébrés, les oligochètes et les poissons (faune). Jusqu'en 2006, sur les réseaux de surveillance nationaux (gestion par le Ministère de l'Ecologie et les Agences de l'Eau), c'est principalement l'indice invertébrés, puis dans une moindre mesure, les indices poissons et diatomées qui étaient utilisés en routine. Depuis 2006, sur les nouveaux réseaux issus de la DCE, l'utilisation de tous les indices biologiques hormis celui des oligochètes se généralise. Les mesures sont effectuées en des lieux précis, de dimensions normalisées et appelés *stations*.

Des données apportant des informations variées sont ainsi disponibles pour évaluer l'état de l'eau au niveau d'une station, mais le problème de leur interprétation demeure, et plus particulièrement la corrélation de l'ensemble des indicateurs. En ce qui concerne les indices biologiques, les experts sont spécialisés dans des groupes taxonomiques particuliers et il leur est donc difficile de proposer une interprétation globale de la qualité de l'eau, intégrant tous les indices. Ainsi croiser et chercher à interpréter l'ensemble des données biologiques, physico-chimiques, physiques, va devenir de plus en plus nécessaire. Cette problématique va être la même quels que soient les pays européens, et ce même si les indices utilisés diffèrent d'un pays à l'autre puisque basés sur des cortèges faunistiques et floristiques différents.

Pour mettre en évidence les influences relatives des indicateurs, les biologistes utilisent classiquement des techniques d'analyse de données telles que l'Analyse des Correspondances Multiples (ACM) (Escofier *et al.*, 1988). L'ACM vise à simplifier un tableau de données mettant en jeu plusieurs variables. Pour cela, de nouvelles dimensions sont calculées sur la base des variables initiales. Chaque dimension se voit associer un pourcentage indiquant la quantité d'informations qu'elle capture. L'intérêt de cette approche est de créer des dimensions capturant la plus grande quantité d'information possible. On étudie alors les données selon ces axes et notamment les plus porteurs d'informations. Pour cela on peut les représenter en considérant les dimensions deux à deux, chaque couple de dimensions permettant de visualiser les données

dans un plan. Ces méthodes reposent donc sur des projections où seulement une partie de l'information est représentée (la somme des quantités portées par les deux axes), ceci limitant les possibilités d'une interprétation globale.

Lafont (Lafont, 2001) a proposé une autre approche, appelée système d'ambiance écologique. La qualité de l'eau y est étudiée selon quatre compartiments (qualité biologique générale, qualité biologique des sédiments, par exemple). Le système identifie les indices biologiques permettant d'évaluer la qualité de chacun de ces compartiments. Enfin, il distingue les différents habitats présents (par exemple les sédiments de surface fins), les associe aux compartiments et ajuste les résultats obtenus pour chaque compartiment en fonction de la surface des habitats et de leur fonctionnement (par exemple la capacité de purification). Ainsi l'habitat "sédiments de surface fins" permet d'évaluer le compartiment "qualité biologique des sédiments", ce compartiment étant associé à l'indice biologique oligochètes. Cette approche ne repose donc pas sur une méthode statistique, mais sur une démarche basée sur les connaissances expertes.

Notre objectif est de développer ce concept et de proposer un outil permettant d'accompagner les biologistes dans l'explicitation de leur expertise et de les aider à mettre en évidence l'ensemble des interactions entre indices biologiques. Nous proposons d'utiliser les treillis de Galois (Barbut *et al.*, 1970), comme outil d'extraction de connaissances sous une forme particulièrement bien adaptée à l'analyse par un expert.

Cet article présente une démarche pour la qualification de stations par les indices biologiques. Il s'agit de définir l'ensemble des profils de stations possibles en se basant sur les concepts mis en évidence par la construction d'un treillis de Galois et l'analyse de ces concepts par un hydrobiologiste. Un profil est la donnée de valeurs possibles pour les indices. L'étude réalisée porte sur un échantillon de stations de la plaine d'Alsace pour lesquelles nous disposons des valeurs concernant quatre indices biologiques (tous les indices sauf les poissons). Cette démarche alterne des étapes d'extraction automatique de connaissances et d'analyse des connaissances extraites.

Nous commencerons par décrire les données traitées. Nous introduirons ensuite les principes des treillis de Galois, puis nous présenterons la démarche proposée pour parvenir à un classifieur de stations. Les premiers résultats obtenus seront discutés avant de conclure.

2. Données

Dans le cadre d'un projet pour l'Agence de l'Eau Rhin-Meuse, le LHyGeS a conçu une base de données destinée à rassembler l'ensemble des informations biologiques, physico-chimiques, chimiques utiles à l'évaluation de la qualité des cours d'eau et à la compréhension des interactions entre les différents paramètres impliqués (Grac *et al.*, 2006). L'étude porte sur un échantillon de stations sélectionnées sur la plaine d'Alsace. La base est alimentée par des données issues de relevés effectués sur ces stations.

Les données utilisées dans cet article concernent 40 stations décrites par quatre attributs :

- l'Indice Biologique Global Normalisé (IBGN), indice basé sur les invertébrés (insectes, mollusques, ...)
- l'Indice Biologique Macrophytique en Rivière (IBMR) – les macrophytes sont des plantes aquatiques visibles à l'œil nu
- l'Indice Oligochètes de Bioindication des Sédiments (IOBS) – les oligochètes sont des vers vivant dans les sédiments
- l'Indice Biologique Diatomées (IBD) – les diatomées sont des algues microscopiques.

Ces indices font partie des cinq indices normalisés par l'AFNOR pour la France (voir par exemple (AFNOR, 2004)). Ils sont établis sur la base des taxons présents dans l'eau et les sédiments. Ces taxons ont des caractéristiques et des exigences particulières, chaque indice apporte des informations spécifiques concernant les quatre compartiments biologiques fondamentaux (Lafont, 2001) : la qualité biologique générale, la qualité biologique des sédiments, la qualité biologique des eaux, la qualité biologique des (communautés de) poissons. Par exemple, l'IOBS apporte des informations sur l'état des sédiments.

Les valeurs pour trois stations sont présentées dans le tableau 1. Chaque valeur est un entier compris entre 1 (très bonne qualité) et 5 (très mauvaise qualité). Un IOBS incalculable indique une présence très faible ou l'absence d'oligochètes.

Code station	IBGN	IBMR	IOBS	IBD
BREI001	3	5	incalculable	3
BRUN002	2	4	3	3
HORG001	3	3	5	3

Tableau 1. Classes de qualité de 3 stations pour 4 indices biologiques

3. Treillis de Galois

Les treillis de Galois ou treillis formels de concepts (Barbut *et al.*, 1970, Davey *et al.*, 1990, Ganter *et al.*, 1997) permettent de traiter des données binaires, spécifiant des propriétés vérifiées ou non par un ensemble d'objets. On appelle *contexte* un triplet (O, T, I) où O est un ensemble d'objets, T un ensemble d'attributs, I une relation de $O \times T$ dans $\{0, 1\}$, et $I(o, t) = 1$ exprime que l'objet $o \in O$ possède l'attribut $t \in T$. Ce contexte se représente naturellement sous la forme d'un tableau binaire.

Par exemple, considérons une version simplifiée de nos données sur les stations. $O = \{S1, S2, S3, S4\}$ est l'ensemble des stations et les attributs de $T = \{\text{Invertébrés, Macrophytes, Oligochètes, Diatomées}\}$ indiquent la présence ou non d'invertébrés, de macrophytes, d'oligochètes et de diatomées (tableau 2).

Station	Invertébrés	Macrophytes	Oligochètes	Diatomées
S1	1	0	0	1
S2	1	1	0	1
S3	1	0	1	1
S4	1	0	1	0

Tableau 2. Exemple de contexte (binaire)

Soient $X \subseteq O$ et $Y \subseteq T$. On définit $f : 2^O \rightarrow 2^T$ tel que $f(X) = \{y \in T \mid \forall x \in X : xIy\}$ est l'ensemble des attributs partagés par tous les objets de X . Par dualité, on définit $g : 2^T \rightarrow 2^O$ tel que $g(Y) = \{x \in O \mid \forall y \in Y : xIy\}$ est l'ensemble des objets qui possèdent tous les attributs de Y . Le couple (f, g) est appelé *correspondance de Galois* entre les ensembles O et T . Les opérations $h = g \circ f$ et $h' = f \circ g$ sont des opérateurs de fermeture, qui sont appelés *fermetures de Galois*.

Ainsi pour notre exemple, si l'on prend $X = \{S1, S2\}$ alors $f(X) = \{\text{Invertébrés}, \text{Diatomées}\}$, ce qui signifie que les stations $S1$ et $S2$ ont en commun la présence d'invertébrés et d'oligochètes. Si maintenant on considère $Y = \{\text{Invertébrés}, \text{Diatomées}\}$, on a $g(Y) = \{S1, S2, S3\}$ car les trois stations $S1$, $S2$ et $S3$ possèdent à la fois les attributs Invertébrés et Diatomées.

Un *concept* du treillis de Galois est un couple (X, Y) , où l'*extension* X est telle que $X \subseteq O$ et $f(X) = Y$ et l'*intension* Y est telle que $Y \subseteq T$ et $g(Y) = X$. Dans notre exemple, $(\{S1, S2, S3\}, \{\text{Invertébrés}, \text{Diatomées}\})$ est un concept du treillis de Galois. Le treillis de Galois est construit à partir de l'ensemble des concepts issus du contexte et d'une relation d'ordre entre ces concepts (inclusion des extensions ou des intensions).

Le treillis de Galois obtenu sur nos données est présenté sur la figure 1. Les concepts sont décrits dans le tableau 3.

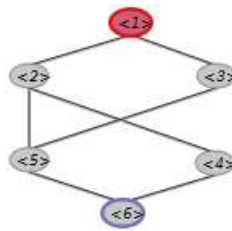


Figure 1. Treillis de Galois du contexte exemple

On peut calculer à partir du treillis la base des règles d'implication entre propriétés (Guigues *et al.*, 1986). On peut également obtenir des règles d'association (Napoli, 2006).

Numéro	Extension (stations)	Intension (indices)
1	{S1,S2,S3,S4}	{Invertébrés}
2	{S1,S2,S3}	{Invertébrés, Diatomées}
3	{S3,S4}	{Invertébrés, Oligochètes}
4	{S2}	{Invertébrés, Macrophytes, Diatomées}
5	{S3}	{Invertébrés, Diatomées, Oligochètes}
6	{}	{Invertébrés, Macrophytes, Diatomées, Oligochètes}

Tableau 3. *Concepts du treillis exemple*

Dans les données sur lesquelles nous travaillons, les attributs peuvent prendre des valeurs dans le domaine {1,2,3,4,5,incalculable}. On parle alors d'attributs multi-valués, et nous nous situons dans le cadre d'un contexte multi-valué tel que défini dans (Ganter *et al.*, 1997). Nous utilisons le logiciel Galicia¹ pour la construction de nos treillis, ce logiciel permettant le traitement de contextes multi-valués.

4. Démarche proposée pour la création du classifieur

Notre but est de pouvoir relier indices biologiques (d'une station) et caractéristiques du milieu telles que le niveau trophique, la qualité des sédiments, les pressions subies, ... Pour cela, nous proposons une démarche conduisant à la création d'un classifieur qui permette de qualifier une station en terme de milieu à partir de ses valeurs pour les indices biologiques. Ce classifieur se basera sur un ensemble de profils de stations déterminés au cours du processus. Un profil est un ensemble cohérent de valeurs possibles pour les indices biologiques, qui correspond à un milieu particulier.

La démarche complète est décrite sur la figure 2. Elle part d'un ensemble de données composées des valeurs des différents indices biologiques pour un échantillon représentatif de stations tel que dans le tableau 1. Ce contexte multi-valué sera appelé *C* dans ce qui suit. La construction d'un treillis de Galois pour *C* fournit des concepts mettant en évidence des stations présentant des valeurs communes pour des indices biologiques. Ces concepts représentent des profils potentiels, et serviront de base pour la détermination des profils définitifs.

L'étape 1 met en œuvre un algorithme de construction de treillis à partir du contexte *C*. Les concepts obtenus, leur organisation hiérarchique dans le treillis et les règles d'implication servent de base à l'analyse de l'expert (étape 2). Le nombre de concepts construits est beaucoup trop important et tous ne représentent pas des profils cohérents. L'analyse de l'expert vise donc à sélectionner les concepts pertinents, notamment en regroupant des valeurs proches pour certains indices. Le regroupement d'indices inclut deux cas : il permet de simplifier le problème et d'identifier des sous-

1. <http://www.iro.umontreal.ca/galicia/>

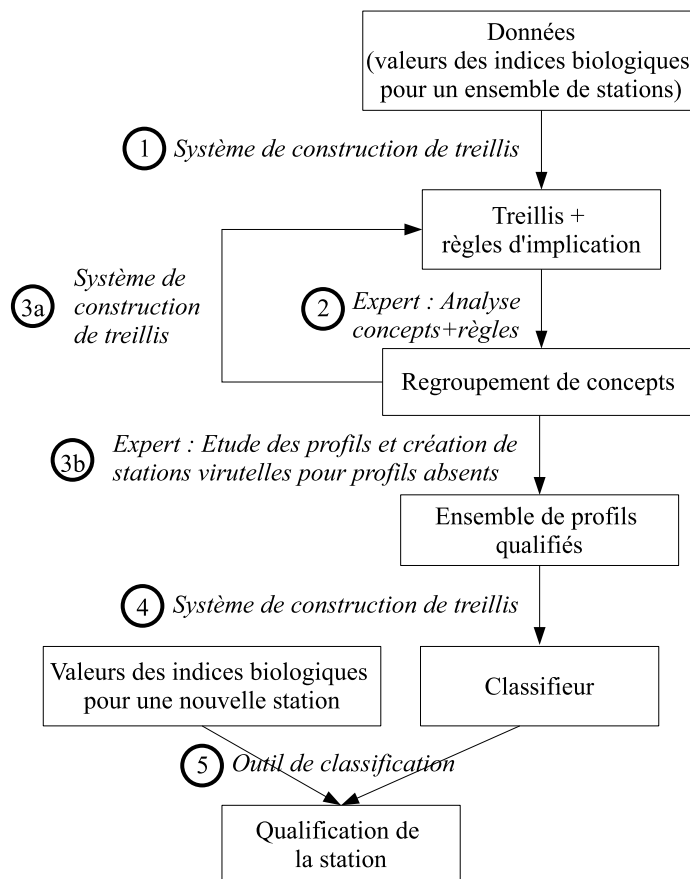


Figure 2. Démarche

treillis basés sur une sélection d'indices (le processus est itératif : étape 3a), ou il permet d'obtenir des profils à qualifier (étape 3b). Dans l'étape 3b, il est ensuite possible à l'expert de déterminer d'éventuels profils non présents dans l'ensemble. Afin de le compléter, l'expert crée des stations virtuelles. Une station virtuelle est décrite par un ensemble de valeurs d'indices correspondant, selon l'expert, aux caractéristiques d'un milieu particulier. L'ensemble des profils étant alors représenté parmi l'ensemble des stations (stations initiales et virtuelles), on construit un classifieur (étape 4) permettant de qualifier une nouvelle station (étape 5).

5. Premiers résultats obtenus sur les données

Le treillis construit à partir des 40 stations (étape 1) comprend 102 concepts, dont sont déduites 51 règles d'implication. Comme on peut le voir sur la figure 3, ce treillis est grand, rendant son interprétation difficile et une simplification serait intéressante.

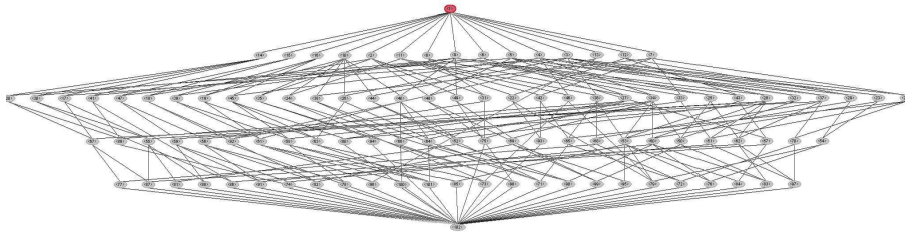


Figure 3. Treillis initial obtenu sur l'ensemble des données

Les règles mettent en valeur des implications entre des valeurs d'indices (pour le jeu de données considéré). Leur examen permet de détecter des corrélations entre indices et des règles similaires ayant des valeurs proches pour des indices. Chaque règle est issue d'un concept et le treillis permet d'identifier facilement les concepts liés (ayant des valeurs communes pour les indices). On peut donc envisager un regroupement des concepts sous-jacents si le concept obtenu est pertinent pour l'expert.

Ainsi dans l'étape 2, l'étude du treillis et des règles a permis d'identifier les règles suivantes, pour lesquelles les valeurs des indices sont identiques à l'exception de l'IBMR (la notation *indice : valeur* indique la valeur pour l'indice) :

Règle	Support
IBMR :3 IOBS :incalculable \Rightarrow IBGN :3 IBD :3	1
IBMR :5 IOBS :incalculable \Rightarrow IBGN :3 IBD :3	2

Une autre implication proche indique : IBMR :4 IOBS :incalculable \Rightarrow IBD :3 (support 6). A cette règle sont associées 6 stations qui ont des IBGN entre 2 et 4. La connaissance qu'a l'expert des stations permet de savoir qu'elles correspondent à des cours d'eau de tailles variées. L'IBD :3 indique une qualité d'eau de niveau assez faible, et l'absence d'oligochètes (IOBS :incalculable) complète cette analyse en indiquant que les sédiments sont en mauvais état.

Parmi les implications, une autre règle présente des caractéristiques similaires : IBGN :3 IBMR :4 IOBS :5 \Rightarrow IBD :3 (support 1). Ainsi, on peut également envisager de regrouper les valeurs d'IOBS 5 et incalculable qui sont contiguës dans l'échelle des valeurs (regroupement accepté par l'expert).

L'étude des règles nous a souvent conduits à former des concepts autour des deux indices IBD et IOBS. Il est donc apparu intéressant de construire un treillis portant uniquement sur ces indices (étape 3a), afin de confirmer les tendances ressorties de cette première analyse et de simplifier l'étude (voir figure 4). D'un point de vue biologique, l'IOBS apporte des indications sur l'état des sédiments et l'IBD sur l'état de l'eau. Le fait de cibler l'analyse sur ces deux compartiments est pertinent pour l'expert. Le treillis, plus petit, comporte 19 concepts, ce qui simplifie son étude.

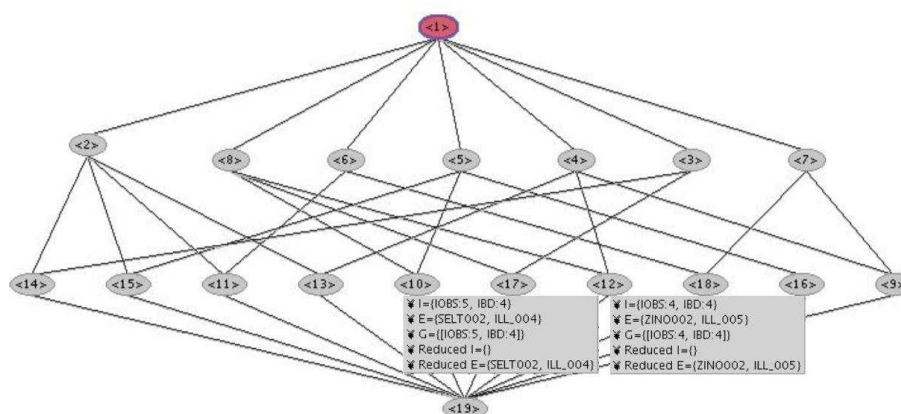


Figure 4. Treillis obtenu à partir des indices IBD et IOBS

Après analyse des concepts et implications de ce nouveau treillis, l'expert a jugé pertinents les regroupements du tableau 4. Les concepts correspondant au regroupement 2 sont indiqués sur le treillis de la figure 4, il s'agit des concepts 10 et 12, et ces deux concepts ont comme père le concept 8 dont l'intension est {IBD :4}.

Numéro	Intension	Nombre de stations concernées
1	IOBS :{incalculable, 5, 4} + IBD :3	22
2	IOBS :{5, 4} + IBD :4	4
3	IOBS :3 + IBD :3	7
4	IOBS :3 + IBD :4	2
5	IOBS :{incalculable, 5, 4} + IBD :2	5

Tableau 4. Regroupements pour le treillis IOBS+IBD

Pour chacun de ces regroupements, nous avons construit un sous-treillis basé sur les stations correspondantes pour ensuite procéder de la même façon (étape 3a). Pour le treillis issu du concept IOBS :{incalculable, 5, 4} + IBD :3, cinq regroupements ont

été identifiés (étape 2) (voir tableau 5). Ces concepts sont en cours de qualification par l'expert (étape 3b), et les sous-treillis issus des quatre autres concepts sont en cours d'analyse (étape 2).

IBMR :{5, 4} + IBGN :2
IBMR :{5, 4} + IBGN :3
IBMR :{5, 4} + IBGN :4
IBMR :3 + IBGN :4
IBMR :2 + IBGN :4

Tableau 5. Regroupements pour le treillis IOBS :{incalculable, 5, 4} + IBD :3

6. Conclusions et perspectives

Nous avons proposé une démarche utilisant les treillis de Galois pour aider les experts dans la détermination de profils de milieux aquatiques décrits par des indices biologiques. Ce travail s'inscrit dans le cadre de la compréhension globale des réponses apportées par les différents indicateurs disponibles pour évaluer la qualité des cours d'eau, l'influence de chacun sur la qualité globale de l'eau n'étant pas à l'heure actuelle connue. L'approche apporte une aide aux experts dans ce travail et offre un support pour la discussion entre experts des différents groupes taxonomiques.

L'analyse faite par l'expert nous a permis de confirmer que les treillis de Galois représentent un outil intéressant pour synthétiser l'information et faciliter sa compréhension. La structuration des concepts donnée par le treillis apporte un plus par rapport à un simple calcul de règles d'association. La démarche proposée a ainsi permis d'assister l'expert dans l'organisation de ses connaissances. L'itérativité du processus, permettant de construire des sous-treillis sur des indices ou valeurs d'indices particuliers, facilite l'analyse.

Les premiers résultats présentés ont été obtenus après analyse d'une hydrobiologiste experte en invertébrés. Nous devons les présenter à des experts des autres indices afin de valider les regroupements de valeurs proposés sur les indices.

L'étape suivante consiste à ajouter des profils manquants (étape 3b) et à construire le classifieur (étape 4), intégrant les profils manquants (stations virtuelles). Différentes approches peuvent être envisagées dont la construction d'un nouveau treillis.

Un autre travail consistera à intégrer aux profils les autres paramètres d'évaluation de l'eau : physico-chimiques et physiques. Pour cela, on pourra repartir des concepts identifiés et construire un treillis où les objets sont les stations et les attributs les concepts, les valeurs de la physico-chimie et les caractéristiques physiques.

Pour 5 stations parmi les 40 étudiées ici, des prélèvements continueront à être effectués régulièrement. La considération de l'évolution des stations au cours du temps

est un aspect important dans le but de comprendre l'influence des pollutions et la réponse des différents indices. Nous comptons aborder cet aspect par la suite.

Ce travail se rapproche de celui proposé par Lafont (Lafont, 2001). Il en diffère car ce dernier propose une évaluation plus fine mais plus complexe, pondérant l'effet de chaque indice en fonction de l'étendue des habitats. Cependant, il ne considère a priori pas le recoupement des résultats obtenus sur plusieurs stations afin de les généraliser. La comparaison des résultats de ces deux approches représente une perspective intéressante.

En comparaison de l'ACM utilisée classiquement par les biologistes, notre approche facilite une interprétation globale, et facilite la création de profils qui sont suggérés par les concepts.

Notons enfin que nous avons considéré l'utilisation de techniques telles que les arbres de décision et de régression, afin de prédire la qualité globale en fonction des réponses des différents indices. Cependant ceci nécessite un étiquetage des différentes stations par la qualité globale. Cet étiquetage global étant justement le but du travail, les techniques supervisées se révèlent non adaptées.

Remerciements

Les auteurs remercient l'Agence de l'Eau Rhin-Meuse pour son soutien à ce projet, ainsi que les relecteurs de leurs conseils pour améliorer la présentation de ce travail.

7. Bibliographie

- AFNOR, « Qualité de l'eau : détermination de l'Indice Biologique Global Normalisé (IBGN) », NF T90-350, 2004.
- Barbut M., Monjardet B., *Ordre et classification – Algèbre et combinatoire*, Hachette, 1970.
- Davey B., Priestley H., *Introduction to Lattices and Order*, Cambridge University Press, 1990.
- Escofier B., Pagès J., *Analyse factorielles simples et multiples*, Dunod, 1988.
- Ganter B., Wille R., *Formal Concept Analysis : Mathematical Foundations*, Springer Verlag, 1997.
- Grac C., Herrmann A., Le Ber F., Trémolières M., Braud A., Handja A., Lachiche N., « Mining a database on Alsatian rivers », in , P. Gourbesville, , J. Cunge, , V. Guinot, , S.-Y. Liang (eds), *Proc. of the 7th Int. Conf. on Hydroinformatics*, vol. III, p. 2263-2270, 2006.
- Guigues J.-L., Duquenne V., « Familles minimales d'implications informatives résultant d'un tableau de données binaires », *Mathématiques et Sciences Humaines*, vol. 95, p. 5-18, 1986.
- Lafont M., « A conceptual approach to the biomonitoring of freshwater : the ecological ambience system », *Journal of Limnology*, vol. 6, p. 17-24, 2001.
- Napoli A., « A smooth introduction to symbolic methods in knowledge discovery », in , H. Cohen, , C. Lefebvre (eds), *Categorization in Cognitive Science*, 2006.

***Atelier « Systèmes d'Information et
de Décision pour l'Environnement »***

26 Mai 2009

Edité par :

Sandro Bimonte, TSCF, Cemagref, Clermont Ferrand

André Miralles, UMR TETIS, Cemagref, Montpellier

François Pinet, TSCF, Cemagref, Clermont Ferrand