
Raisonner à partir de cas d'allocation dans des parcelles agricoles

Florence Le Ber¹, Xavier Dolques¹, Laura Martin², Alain Mille³, Marc Benoît²

1. Université de Strasbourg, ENGEES, CNRS, ICube UMR 7357,
F-67400 Illkirch-Graffenstaden, France

florence.leber@engees.unistra.fr, xavier.dolques@laposte.net

2. SAD ASTER, UR 055, INRA Mirecourt, France

marc.benoit@mirecourt.inra.fr

3. Liris, Université Lyon1, UMR CNRS 5205, Lyon, France

alain.mille@univ-lyon1.fr

RÉSUMÉ. Cet article présente un prototype de raisonnement à partir de cas développé dans le cadre d'une application en agronomie, pour prévoir l'implantation dans les exploitations agricoles d'une nouvelle culture énergétique, le miscanthus. Des enquêtes ont été réalisées pour accéder au mode de décision des agriculteurs. Néanmoins, peu d'expériences étaient disponibles pour construire ce modèle, ce qui nous a conduits à intégrer des variantes dans le module de raisonnement, permettant ainsi des usages différents. Nous présentons quelques résultats expérimentaux, puis discutons du modèle et de ses usages.

ABSTRACT. This article describes a prototype of case-based reasoning, built for the agronomical domain. Its aim is to forecast the location of a new energy crop, the miscanthus. Interviews were conducted with farmers in order to know how they make their decisions. Nevertheless, few experiences were available for building this model, and we therefore introduced variations in the reasoning modules, for allowing various uses. Experimental results are presented. The model and its possible uses are then discussed.

MOTS-CLÉS : raisonnement à partir de cas, adaptation, explication, culture énergétique.

KEYWORDS: case based reasoning, adaptation, explanation, energy crop.

DOI:10.3166/RIA.31.681-707 © 2017 Lavoisier

1. Introduction

Parmi les modèles à base de connaissances, le raisonnement à partir de cas (RÀPC), introduit par (Riesbeck, Schank, 1989), permet d'appréhender des domaines où les connaissances générales sont mal formalisées et où l'expertise repose majoritairement sur l'expérience. Parmi ces domaines on peut citer la médecine, la chimie, l'ingénierie, la prévention des risques ou encore, plus prosaïquement, la cuisine, où des systèmes à partir de cas ont pu être développés avec succès (voir par exemple (Hammond, 1990; Herbeaux, Mille, 1999; Avesani *et al.*, 2000)). Classiquement le RÀPC est vu comme un mode de résolution de problèmes s'appuyant sur la réutilisation de solutions de problèmes déjà résolus. Un *cas* est la représentation d'un épisode de résolution de problème : il encode un problème et une solution de ce problème ainsi que des informations sur les liens entre ce problème et cette solution. L'ensemble des cas d'un système de RÀPC est regroupé dans une *base de cas* et un cas de cette base est appelé *cas source* (dont la partie problème est appelée *problème source*).

Raisonnement à partir de cas, c'est résoudre un problème, appelé *problème cible*, en s'appuyant sur la base de cas et sur des connaissances du domaine d'application. Ce raisonnement s'effectue en deux étapes principales, la *remémoration* et l'*adaptation*. La remémoration consiste à sélectionner un cas source jugé « similaire » au problème cible. L'adaptation utilise le cas source remémoré pour résoudre le problème cible, souvent en modifiant la solution associée au cas source et en s'appuyant sur la différence entre le problème source et le problème cible. Cependant, dans certains domaines, le modèle générique de cas, problème-solution, exigeant une forme fixe aussi bien pour la partie problème que pour la partie solution, n'est pas directement opérationnel. Il faut alors introduire une étape d'élaboration du problème cible, comme présenté par (Fuchs *et al.*, 2006), permettant de reformuler les choses afin de pouvoir mobiliser les connaissances effectivement disponibles pour l'adaptation d'une solution existante.

La construction d'un système de RÀPC nécessite un travail préalable de modélisation de connaissances. Cette étape peut souvent s'appuyer sur des documents plus ou moins normalisés, comme un dossier patient, ou une recette de cuisine (voir des exemples dans (Brüninghaus, Ashley, 2001)). Les techniques de fouille de données ou d'analyse de textes peuvent alors être employées (Adeyanju *et al.*, 2009; Dufour-Lussier *et al.*, 2014). Toutefois, dans de nombreuses applications du RÀPC, la définition et la modélisation des cas et des connaissances nécessaires aux étapes de remémoration et d'adaptation requièrent un long travail d'analyse commun entre experts du domaine et informaticiens (voir par exemple l'expérience relatée par (Le Ber, Brassac, 2008)). La notion de cas ne s'applique alors pas de la même façon, mais garde sa pertinence si on considère que l'on cherche à expliquer une situation à partir de l'explication trouvée dans une situation similaire : il s'agit d'une approche formalisée par (Riesbeck, Schank, 1989) et qui permet de prendre en compte des domaines en cours d'exploration, comme le domaine qui nous préoccupe ici, ou relevant de l'argumentation, comme en jurisprudence (Bain, 1989) ou en enseignement (Burke, Kass, 1996).

Le domaine qui nous préoccupe ici est le domaine agronomique, dans une situation où même l'expérience est rare, puisqu'il s'agit de l'introduction dans des exploitations agricoles d'une nouvelle culture, le miscanthus, pour laquelle il n'existe pas encore de référentiel agronomique. Nous avons développé une démarche et un outil de RÀPC qui s'appliquent donc à une expérience en construction. Les questions à résoudre sont : quel est le processus de décision qui conduit un agriculteur à implanter du miscanthus dans une parcelle ? ce processus est-il généralisable ou adaptable pour prévoir ce que fera un autre agriculteur sur une autre parcelle ? Pour répondre à ces questions, un prototype, dénommé SAMM¹, a été réalisé et expérimenté en exploitant les connaissances collectées à partir d'enquêtes menées en Bourgogne, dans une zone où l'exploitation du miscanthus s'est développée il y a peu, sous l'impulsion d'une coopérative (Martin, 2014). Ces connaissances sont de deux types, d'une part des connaissances sur les caractéristiques des parcelles et, d'autre part, des explications données par les agriculteurs sur les raisons de leur choix d'implanter ou non du miscanthus dans une parcelle. Le choix du raisonnement à partir de cas est le pendant d'une approche par apprentissage automatique, expérimentée par ailleurs (Rizzo *et al.*, 2014); le RÀPC permet, à notre sens, de mieux explorer les modes de raisonnement et les critères de choix des agriculteurs (Martin, Le Ber, 2013). L'utilisateur ou utilisatrice final visé est un décideur, en charge d'une coopérative agricole ou d'une unité de production de chaleur, qui, dans un cadre prospectif, voudrait estimer la capacité d'un territoire à produire du miscanthus en tenant compte des critères propres aux agriculteurs.

Des versions préliminaires ou plus concises de ce travail ont été présentées dans (Le Ber *et al.*, 2015 ; 2017). Nous le décrivons ici de façon détaillée à la fois sous les aspects applicatifs, concernant les connaissances et données agronomiques, et sous les aspects informatiques, concernant le modèle construit et les expérimentations menées. Le plan de l'article est le suivant : après un état de l'art (section 2), nous exposons la problématique agronomique et la procédure d'acquisition de connaissances mise en place (section 3); nous détaillons ensuite le modèle SAMM (section 4) et en donnons une évaluation (section 5), puis en discutons (section 6) avant de conclure.

2. Travaux connexes

Notre travail s'inscrit dans les travaux liant intelligence artificielle et agronomie (Guyet *et al.*, 2013). Plus précisément, il s'inspire des travaux de (Osty *et al.*, 2008) qui ont discuté de l'intérêt du RÀPC comme outil de modélisation pour l'agronomie des territoires (Benoît *et al.*, 2012). Cette discussion s'appuyait sur une expérience de développement d'un système pour l'analyse et la comparaison d'enquêtes en exploitations (Le Ber *et al.*, 2003). Ce système utilisait des graphes conceptuels, dont les sommets et les arêtes sont étiquetés, pour représenter l'organisation spatiale de territoires agricoles. Le raisonnement combinait la classification hiérarchique au sens des logiques de description et la composition de relations spatiales qualitatives. Plus lar-

1. Pour *Spatial Allocation Modelling of Miscanthus*.

gement, il existe de nombreux systèmes liant RÀPC et sciences environnementales, la plupart adoptant des approches très numériques, proches de méthodes d'apprentissage. On citera en particulier (Du *et al.*, 2012), qui intègrent des relations spatiales entre voisins pour mesurer la similarité entre parcelles et prédire leur occupation du sol (bâti, forêt ou culture). Des travaux plus anciens avaient déjà utilisé le RÀPC pour l'analyse de données géographiques, par exemple pour la classification des sols (Holt, Benwell, 1996).

Dans les travaux fondés sur des approches numériques, il est peu fait mention des connaissances mises en jeu par les acteurs, comme nous avons entrepris de le faire ici. Le système décrit par (Le Ber *et al.*, 2003) se fondait sur des connaissances d'acteurs préformalisées par des chercheurs. (Girard *et al.*, 2001) ont pris en compte des connaissances d'acteurs dans une approche de modélisation proche du RÀPC, mais sans aller au développement d'un système, du fait de la complexité des formes de connaissances à modéliser (conduites de troupeaux ovins). De fait, les systèmes incorporant explicitement des expériences d'acteurs afin de les partager, comme celui décrit par (Bosch *et al.*, 1997) sur les *rangeland* neo-zélandais, sont rares. Citons encore les travaux autour du système CARMA pour le diagnostic et le traitement d'invasions de ravageurs des cultures (Hastings *et al.*, 2010), qui adapte des modèles établis sur des connaissances expertes, et a pu être généralisé à plusieurs états américains.

Le travail présenté ici s'inscrit dans une situation d'utilisation de l'expérience, proche du *Explanation Based Reasoning* (Schank *et al.*, 1994). L'approche d'un raisonnement analogique fondé sur les explications est bien documenté dans la littérature avec plusieurs synthèses toutes produites dans les années 2004-2005. (Roth-Berghofer, 2004) reprend les fondements de la notion d'explication aussi bien au niveau philosophique que dans le cadre d'un système à base de connaissances. Il distingue les explications cherchant à répondre aux questions suivantes : quel est le concept (qui intervient dans la décision) ? quelle est la justification d'un fait ? comment une décision s'est-elle mise en place ? dans quel but telle décision a-t-elle été produite ? quel est le contexte cognitif de telle décision ? Les auteurs énumèrent quelques bonnes propriétés de *bonnes* décisions : fidélité, suffisance, facilité d'élaboration, efficacité. Dans une autre synthèse, (Sørmo *et al.*, 2005) s'intéressent spécifiquement à la capacité d'un système à base de connaissances à fournir les explications permettant à un utilisateur ou une utilisatrice de s'approprier le raisonnement et ses conclusions. Le RÀPC se prête particulièrement bien à cet exercice en gardant explicitement les explications spécifiques à chaque cas. Un cas, exprimé dans le registre de l'utilisateur, avec des explications faisant appel à des concepts partagés, peut être plus facilement réutilisé pour en résoudre un autre tout en fournissant les clés de compréhension du domaine. L'ouvrage préfacé par (Leake, McSherry, 2005) reprend ces différentes façons de considérer, représenter et raisonner sur les explications.

3. Problématique

3.1. *Le miscanthus, une culture énergétique pérenne*

Le *Miscanthus x giganteus* – dit miscanthus – est une plante herbacée originaire d'Asie, de grande taille (voir figure 1), cultivée en France depuis 2006 (Schnitzler, 2011). Cette plante est pérenne, elle est cultivée pendant 15 à 20 ans sur une même parcelle et peut produire une biomasse importante tout au long de cette période. Elle est cultivée pour des usages exclusivement non alimentaires : sa biomasse est utilisée comme écomatériau, comme combustible pour produire de la chaleur et pour le paillage.



Figure 1. *Vue d'une parcelle de miscanthus*

L'extension de l'implantation du miscanthus en France est importante et peut constituer une tendance lourde au vue de l'intérêt récent des dirigeants politiques. Cet intérêt s'explique en partie par son potentiel pressenti en termes de bilan énergétique et de bilan environnemental : le miscanthus nécessite peu d'apports, fertilisants ou produits de traitement, et peu d'interventions de l'agriculteur. En revanche le miscanthus a des exigences en matière de substrat, il se développe préférentiellement dans les sols profonds et humides et peut se trouver en concurrence avec d'autres cultures, comme le maïs. Sa pérennité a deux conséquences : les choix effectués par l'agriculteur l'engagent pour de nombreuses années, d'où une réflexion attentive avant l'implantation ; la conduite de la parcelle est alors durablement simplifiée, le miscanthus ne nécessitant après son implantation qu'une récolte annuelle de la biomasse séchée sur pied en fin d'hiver. Ainsi, modéliser la localisation du miscanthus dans les territoires permet d'identifier et d'anticiper les risques d'une concurrence d'usage des sols et plus globalement de la réorganisation des territoires, en cas de forte demande énergétique.

Les connaissances sur les processus d'allocation du miscanthus sont encore peu nombreuses. Les travaux existants portent principalement sur les facteurs d'adoption

du miscanthus dans les exploitations agricoles, en particulier sur les facteurs économiques. Plus récemment d'autres facteurs ont été étudiés, comme la taille de l'exploitation agricole, le capital humain (âge, éducation, valeurs morales et sociales), les contraintes techniques ou en ressources, l'accès à l'information et la confiance (Bocquého, 2012). Or modéliser la localisation du miscanthus requiert de prendre en compte les choix des agriculteurs et de coupler des variables biophysiques et humaines. Il est nécessaire pour cela de recueillir et d'analyser un grand nombre de données et d'informations plus ou moins complexes, en exploitant des sources de données et des enquêtes auprès des exploitants concernés.

3.2. Des enquêtes pour collecter les connaissances

Des enquêtes ont été réalisées en 2011 puis renouvelées l'année suivante auprès de treize exploitations agricoles dans une zone d'implantation du miscanthus, en Côte d'Or (Bourgogne). Pour établir les règles d'allocation du miscanthus, il a été demandé aux enquêtés de décrire leur exploitation îlot par îlot², en explicitant leurs décisions d'allocation de culture. Nous disposons ainsi de la description fine des îlots et des objectifs de l'enquêté. Cette description permet de collecter des critères, relatifs à chacun des agriculteurs enquêtés, et nécessaires à prendre en compte pour modéliser l'allocation du miscanthus dans le parcellaire des agriculteurs (Martin *et al.*, 2012).

Nous recherchons aussi les règles d'allocation spatiale du miscanthus choisies par les agriculteurs dans la situation réelle et dans deux mises en situation imaginées : 1) les îlots réellement implantés en miscanthus n'existent pas – il est alors demandé à l'enquêté où le miscanthus aurait été implanté ; 2) l'enquêté souhaiterait planter d'autres îlots en miscanthus – il lui est alors demandé dans quels îlots supplémentaires et sous quelles conditions. Ces deux mises en situation permettent de construire des cas et des ensembles de règles. Par exemple, un agriculteur a implanté du miscanthus sur une parcelle A, mais s'il voulait planter davantage de miscanthus (hypothèse 2) alors il l'aurait fait sur une parcelle B. Cette parcelle B est donc *implantable en miscanthus sous conditions* (par exemple, si la filière s'avérait pérenne ou si le revenu espéré était plus élevé).

Après dépouillement des enquêtes et retranscription des entretiens, nous avons constitué un ensemble d'informations et connaissances qui se présentent synthétiquement de la façon suivante :

- une liste de parcelles dans lesquelles a été, pourrait être implanté ou non du miscanthus,
- où chaque parcelle est décrite par des caractéristiques de différents types,
- et assortie d'un ou plusieurs énoncés exprimant les raisons du choix de l'agriculteur vis-à-vis du miscanthus.

2. Un îlot est un groupe de parcelles voisines gérées de la même manière.

Par exemple dans l'énoncé suivant (1) « *Une parcelle pas drainée : très mauvaise, enfin humide, humide, très humide (...) j'ai tout le temps vu en jachère (...) ce n'est pas le même prix : je dis « hop, je fais le miscanthus dedans »* », on voit que la décision d'implanter prise par l'agriculteur se fonde sur le caractère humide de la parcelle et sur son occupation précédente par une jachère peu rémunératrice. Dans un énoncé tiré d'un autre entretien (2) « *Celle-là elle devait y être mais elle n'y est pas car je l'ai mise en bio. Trop loin, pas de forme, très hydromorphe, c'est un de mes plus mauvais champs. Et puis quand j'ai vu qu'ils me faisaient suer avec leurs plants pourris, j'ai dit « allez stop, on arrête là le miscanthus » et puis comme j'ai vu que je passais en bio, j'ai mis ce champ en bio* », on remarque que malgré le caractère hydromorphe de la parcelle, qui est aussi éloignée et biscornue, donc difficile à cultiver, l'agriculteur ne fait pas le choix du miscanthus, du fait de sa stratégie générale.

3.3. Modélisation du domaine

Les différents arguments utilisés par les agriculteurs renvoient à des caractéristiques des parcelles (régime hydrique, forme, taille, etc.) qui ont été recensées et transformées en éléments de la forme attribut-valeur (voir section 4.1). Chaque attribut-valeur est assorti d'une influence sur le potentiel d'implantation de la parcelle pour le miscanthus. Cette influence est estimée en décomptant, dans l'ensemble des entretiens, le nombre de fois où la caractéristique est utilisée positivement ou négativement dans les choix des agriculteurs. Une influence positive correspond à une caractéristique citée majoritairement par des agriculteurs ayant choisi d'implanter du miscanthus sur une parcelle. Inversement, une influence négative correspond à une caractéristique citée majoritairement par des agriculteurs ayant choisi de ne pas implanter du miscanthus sur une parcelle. Si l'argument est utilisé équitablement alors l'influence est neutre. Par exemple, l'attribut-valeur (taille, petite) a une influence positive, tandis qu'à l'inverse, l'attribut-valeur (taille, grande) a une influence négative.

Dans l'approche développée ici, les attributs sont considérés comme indépendants d'un point de vue informationnel (des prétraitements ont permis d'élaguer les attributs portant éventuellement la même information). En revanche, une certaine interdépendance vis-à-vis de la décision existe : il ressort du discours des agriculteurs un poids plus fort ou un effet combiné (par exemple une parcelle de forme étroite, avec un voisinage arboré, voir les exemples détaillés présentés en section 6) de certains attributs qui interviennent dans leur processus de décision ; mais les combinaisons et les poids sont variables selon les agriculteurs, ce qui rend difficile la mise au jour de régularités, qui pourraient ensuite être prises en compte par le cycle du RÀPC, dans les mesures de similarité et les procédures d'adaptation.

4. Le modèle SAMM

Comme annoncé en introduction, le système doit permettre de prévoir si une parcelle d'un agriculteur est implantable ou non en miscanthus, ou encore si cette parcelle est implantable sous conditions. Le prototype construit dans ce but, intitulé SAMM,

est constitué d'une base de cas, d'une base de règles, d'une base de connaissances (connaissances de similarité et d'adaptation) et d'un module de raisonnement qui sont décrits ci-après.

4.1. Base de cas

Dans notre application, un cas est défini comme une expérience particulière d'insertion territoriale du miscanthus décidée par un agriculteur. Le couple problème-solution est une parcelle agricole et son potentiel d'implantation du miscanthus (voir figure 2). Chaque cas est représenté par un vecteur d'attributs-valeurs, composé de deux parties :

1. la partie problème décrit les caractéristiques de la parcelle ; les attributs sont choisis parmi 32 attributs influençant le potentiel d'implantation du miscanthus (voir tableau 1), qui relèvent de six catégories (géométrique, d'accessibilité, etc.) et sont décrits par 159 valeurs ; le sous-ensemble des descripteurs du problème est appelé DP (pour *descripteurs du problème*) et se présente formellement comme un ensemble de couples $(a, v) \in \mathcal{A} \times \mathcal{V}$, où \mathcal{A} et \mathcal{V} sont respectivement l'ensemble des attributs et l'ensemble des valeurs ;

2. la partie solution décrit le potentiel d'implantation pour le miscanthus par une seule variable prenant trois valeurs : non implantable en miscanthus (valeur 0), implantable en miscanthus (valeur 1) et implantable sous conditions (valeur 2) ; la solution s'écrit donc comme un couple (miscanthus, i) où $i = 0, 1$ ou 2 .

Au cas est également associé un identifiant représentant l'agriculteur exploitant la parcelle.

La base de cas du prototype SAMM comprend 82 parcelles pour lesquelles le potentiel d'implantation pour le miscanthus a été explicité par les agriculteurs lors des enquêtes en exploitations agricoles ; il s'agit donc de parcelles pour lesquelles la solution est connue. L'organisation de cette base de cas est plate, au sens où tous les cas sont considérés au même niveau ; toutefois, pour faciliter les étapes de remémoration et d'adaptation, des index sont utilisés. Chaque cas est donc affecté d'un sous-jeu de descripteurs, appelé DI (pour *descripteurs de l'index*) qui correspond aux éléments qui ont effectivement pesé dans la décision de l'agriculteur. Dans l'exemple (1) cité ci-dessus, le régime hydrique de la parcelle et son occupation précédente vont former les descripteurs de l'index.

Deux exemples de cas sont décrits dans le tableau 2. Seuls sont représentés les attributs valués pour ces cas, c'est-à-dire les attributs qui ont été cités par l'agriculteur et qui sont donc supposés intervenir dans sa prise de décision. On ne connaît pas la valeur des autres attributs pour ces parcelles. On remarque que les descripteurs du problème (DP) sont peu nombreux (6 ou 7 pour 32 possibles) et que les descripteurs de l'index (DI) utilisés explicitement pour la décision le sont encore moins (ici 2 ou 3, marqués par une croix). Tous les attributs sont discrets car la description des cas est issue des entretiens avec les agriculteurs et s'appuie donc sur des données verbales, nominales.

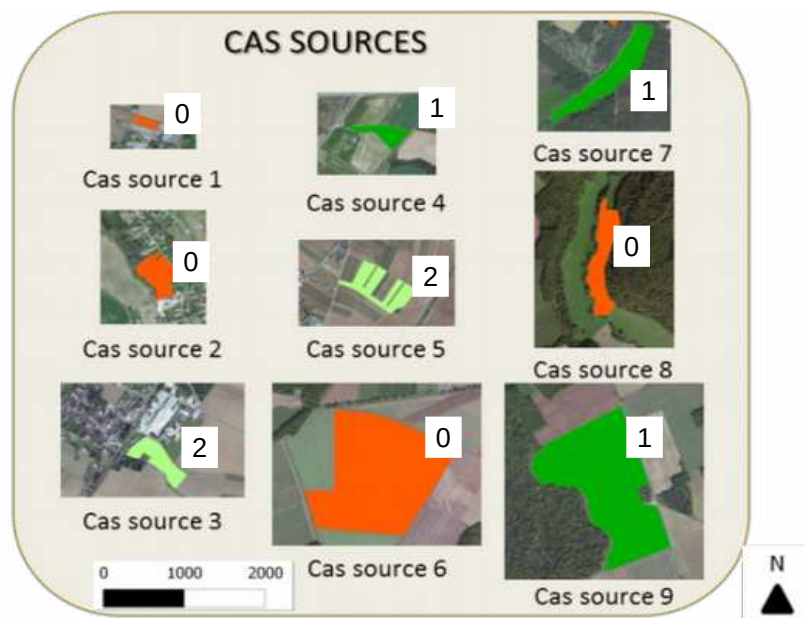


Figure 2. Géovisualisation de quelques cas sources : les parcelles sont colorées selon leur potentiel d'implantation (0/rouge, 1/vert, 2/vert clair) (Martin, 2014)

La difficulté – non traitée dans cet article mais abordée dans l'article (Rizzo *et al.*, 2014) – est de mettre en relation ces attributs nominaux avec des données continues, par exemple issues d'un système d'information géographique, pour construire les problèmes cibles. Certains informations peuvent être difficiles à obtenir (par exemple la localisation du siège d'exploitation) mais d'autres sont couramment utilisées dans les applications agronomiques (occupation précédente, géométrie des parcelles, type de sol, régime hydrique ...).

4.2. Base de règles

Le système SAMM dispose également d'un ensemble de règles qui formalisent les explications données par chaque agriculteur concernant ses modes de décision. Nous les appelons règles de décision. Elles sont de deux types : les règles de décision génériques et non situées dans l'espace (dites aspatiales) et les règles de décision liées à une parcelle, dites spatialisées. Les règles aspatiales traitent par exemple des conditions économiques ou relatives à la réglementation environnementale tandis que les règles spatiales traitent des caractéristiques propres des parcelles.

Les règles prennent la forme générale $\langle \bigwedge \text{prémises (descripteurs)}, \text{conclusion (potentiel d'insertion)} \rangle$. Voici quelques exemples de règles spatiales :

R1 : $\langle (\text{excès d'eau} = \text{inondable}), (0) \rangle$

Tableau 1. Exemples de quelques descripteurs, attributs et valeurs possibles

Attribut	Valeur	Attribut	Valeur		
occupation précédente	culture fourragère culture pérenne jachère prairie rotation culturale	forme de la parcelle	assez large assez longue assez longue ET assez large assez longue ET étroite assez longue ET pointue ET convexe assez régulière contraignante convexe étroite irrégulière irrégulière ET assez longue irrégulière ET étroite partiellement irrégulière partiellement peu longue peu longue pointue pointue ET convexe pointue ET convexe ET étroite pointue ET convexe ET peu longue pointue ET étroite régulière ET assez longue très convexe régulière		
potentiel agronomique	bon mauvais moyen très bon très mauvais		voisinage cultures	vignes miscanthus prairie verger	
régime hydrique du sol	humide hydromorphe non humide non hydromorphe non séchant partiellement hydromorphe partiellement séchant partiellement très humide partiellement très séchant séchant séchant ET humide séchant ET hydromorphe séchant ET très humide très humide très séchant très séchant ET humide un peu humide un peu séchant		protection environnement	natura 2000 périmètre captage eau potable zone vulnérable	
	statut foncier		location intégrale propriété partielle propriété partielle propriété intégrale	voisinage arbres	arbres isolés arbres isolés ET bois bois bois ET haie bosquet bosquet ET bois forêt forêt ET bosquet forêt ET haie haie lande
distance au siège exploitation	loin loin ET isolé loin ET non isolé proche très loin très proche un peu loin		largeur d'accès		étroit très étroit

R2 : < (taille = moyen), (2) >

R3 : < (drainage = inexistant) \wedge (excès d'eau = résurgences et sources), (0) >

R4 : < (occupation précédente = jachère) \wedge (régime hydrique = humide), (1) >

Chaque règle est étiquetée par un identifiant associé à l'agriculteur qui l'a énoncée, et par là rattachée aux cas sources étiquetés par le même identifiant. Notons que d'un agriculteur à l'autre, les règles peuvent être incohérentes, puisque, selon l'individu, les caractéristiques des parcelles peuvent avoir des influences positives ou négatives sur le

Tableau 2. Deux exemples de cas – DP : descripteurs du problème ; DI : descripteurs de l'index (marqués par x)

Attribut	Source 1		Source 2	
	DP	DI	DP	DI
occupation précédente	jachère	x	jachère	x
potentiel agronomique	moyen		mauvais	x
régime hydrique	résurgences		inondable	
taille de la parcelle	ETsources			
distance au siège	petite	x		
largeur obstacle	proche		loin	x
franchissement	étroit			
largeur franchissement			pont	
statut foncier	propriété		étroit	
Potentiel d'implantation	1		1	

choix d'implantation. Par exemple, pour un agriculteur A_1 , on a la règle \langle (distance au siège = loin), (2) \rangle et, pour un agriculteur A_2 , la règle \langle (distance au siège = loin), (0) \rangle .

La base de règles recouvrant tous les agriculteurs est constituée de 96 règles : 61 règles concluent sur 0 (parcelle non implantable en miscanthus), 8 règles concluent sur 1 (parcelle implantable) et 27 règles concluent sur 2 (parcelle implantable sous conditions). Le faible nombre de règles concluant sur le caractère implantable d'une parcelle met en exergue le caractère novateur et encore rare du miscanthus. C'est pourquoi dans la suite nous nous intéresserons essentiellement au fait qu'une parcelle soit implantable sous conditions ou non implantable.

4.3. Connaissances de similarité et d'adaptation

Les connaissances de similarité ont été élaborées à partir des résultats des enquêtes sur les décisions des agriculteurs. Différents niveaux et types de similarité sont considérés (dans la suite l'indice s fait référence au cas *source*, l'indice c au cas *cible*) :

– un niveau numérique global : le nombre de descripteurs partagés (quelle que soit la valeur) entre les problèmes *source* et *cible*, noté n ; on peut considérer les ensembles DP_s du cas *source* et DP_c du problème *cible*, alors :

$$n_{DP}(s, c) = |\{a \in \mathcal{A} | \exists(a, v_1) \in DP_s \wedge \exists(a, v_2) \in DP_c\}|$$

ou bien les ensembles DI_s et DP_c (l'ensemble DI pour *cible* n'est pas connu), et alors :

$$n_{DI}(s, c) = |\{a \in \mathcal{A} | \exists(a, v_1) \in DI_s \wedge \exists(a, v_2) \in DP_c\}|$$

– un niveau sémantique local : les valeurs des descripteurs communs sont comparées deux à deux ; si elles sont égales, la distance locale (notée d_l) est nulle ; si elles sont différentes, la distance varie entre 5 et 20, selon qu'elles influencent de la même façon ou non le potentiel d'implantation ; par exemple la valeur de régime hydrique "terre humide" est similaire à la valeur "résurgences et sources", mais distante de la valeur "terres séchantes" ; plus précisément :

- $d_l((a, v_s), (a, v_c)) = 5$ si les deux attributs-valeurs ont des influences similaires, positives ou négatives, quant à l'implantation du miscanthus ;

- $d_l = 20$ si les deux attributs-valeurs ont des influences différentes, l'une positive et l'autre négative ;

- $d_l = 10$ si au moins un des attributs-valeurs a une influence neutre.

Ces seuils de distance ont été choisis de façon heuristique pour tenir compte des différences de valeurs en distinguant suffisamment les attributs dont les valeurs ont la même influence de ceux où les valeurs ont des influences opposées.

– un niveau sémantique global : la distance globale entre source et cible est la moyenne des mesures locales sur les attributs communs ; elle peut être calculée pour $n = n_{DP}$ ou $n = n_{DI}$:

$$d_g(s, c) = \frac{\sum_{i=1}^n d_l((a_i, v_{is}), (a_i, v_{ic}))}{n}$$

Dans les systèmes de RÀPC existants, l'adaptation s'opère quand au moins un descripteur du problème source est dissemblable du problème cible. Les connaissances utilisées pour l'adaptation s'appuient sur les relations de dépendance entre problèmes et solutions, relations dont l'acquisition peut être manuelle ou automatique. Ces connaissances permettent alors de modifier la solution du cas source sélectionné pour construire une solution adaptée au problème cible.

Pour construire une telle solution, nous utilisons ici les règles de décision des agriculteurs. L'idée est soit de recopier la solution du cas source, soit d'utiliser les règles de l'agriculteur associé au cas source (ou d'un ensemble d'agriculteurs) pour construire cette solution. Les connaissances d'adaptation permettent, parmi plusieurs règles applicables, de choisir celle qui donnera la solution. Ce choix s'appuie sur un contexte d'adaptation : l'utilisateur peut choisir de favoriser les règles concluant sur 0 (si le contexte n'est pas favorable au miscanthus, par exemple parce que son prix est faible par rapport aux cultures traditionnelles) ou au contraire celles qui concluent sur 2 ou 1, si le contexte est favorable au miscanthus.

4.4. Module de raisonnement

Le raisonnement dans le prototype SAMM est constitué de deux étapes principales, la remémoration de cas sources similaires à un problème cible et l'adaptation d'un cas source au problème cible. Préalablement à l'adaptation, la sélection d'un cas source est réalisée.

4.4.1. Remémorer des cas sources

La remémoration consiste à identifier un ou plusieurs cas sources de la base de cas susceptibles d'aider à la résolution du problème cible. Elle se déroule en trois temps : l'appariement du problème cible aux problèmes sources, l'évaluation de la similarité entre les problèmes et la sélection finale des cas sources. Dans le prototype SAMM, l'appariement du problème cible aux cas sources consiste en un appariement de vecteurs. L'évaluation de la similarité entre chaque problème source et le problème cible s'appuie sur les mesures définies ci-dessus. La sélection est réalisée selon différentes combinaisons de ces mesures (cf. figure 3).

Quatre algorithmes de remémoration ont été définis. Ils se caractérisent par les descripteurs utilisés pour la comparaison des problèmes et la procédure de remémoration choisie :

- il y a deux ensembles de descripteurs : le premier ensemble, DP, est composé des descripteurs du problème source ; le second, DI, est composé des descripteurs qui indexent le cas source (cf. paragraphe 4.3) ;
- il y a deux procédures : (i) la première filtre d'abord les cas sources minimisant la distance globale puis les cas sources maximisant le nombre de descripteurs comparés et (ii) la seconde, à l'inverse, filtre dans un premier temps les cas sources maximisant le nombre de descripteurs comparés puis les cas sources minimisant la distance globale.

Dans la version prototype du système, le filtrage consiste à sélectionner le cas minimisant la distance ou maximisant le nombre de descripteurs. La deuxième étape n'est appelée que si la première renvoie plusieurs cas sources. Dans une version ultérieure, le filtrage peut s'envisager selon un seuil à choisir par l'utilisateur.

```

Comparaison cas source 10499951 avec cas cible 10385231
distance pb entre odsP_jachere et odsP_jachere : 0
distance pb entre hydr_sechant et hydr_partiellement hydromorphe : 20
distance pb entre forme_pointue et forme_pointue ET convexe : 10
distance index entre odsP_jachere et odsP_jachere : 0
distance index entre forme_pointue et forme_pointue ET convexe : 10
nombre d'éléments pb comparés : 3
distance globale pb : 10,0
nombre d'éléments index comparés : 2
distance globale index : 5,0

```

Figure 3. Exemple de résultats obtenus après comparaison d'un cas source à un problème cible : la similarité est calculée sur DP puis sur DI

Dans la suite nous désignons les quatre algorithmes de remémoration respectivement par : RM1 (DP et (i)), RM2 (DP et (ii)), RM3 (DI et (i)), RM4 (DI et (ii)).

4.4.2. Adapter la solution des cas sources

Dans le prototype SAMM, l'adaptation est réalisée par substitution (recopie) de la solution source ou par transformation de cette solution en utilisant les règles décrites

ci-dessus. Le résultat de la remémoration permet de proposer une solution pour le problème cible par simple recopie. Si plusieurs cas sources ont été remémorés, la solution est déterminée par un vote majoritaire. Si aucune majorité ne se dégage on garde une solution complexe parmi les combinaisons 0-1, 1-2, 0-2, 0-1-2.

Pour la transformation, plusieurs ensembles de règles peuvent être considérés : soit l'ensemble des règles associées à l'agriculteur du cas source remémoré, soit l'ensemble des règles associées à un sous-groupe d'agriculteurs (par exemple les agriculteurs d'une même petite région que l'agriculteur associé au cas source), soit l'ensemble de toutes les règles. Les règles sont applicables dès lors que leurs prémisses sont similaires aux descripteurs du problème cible, c'est-à-dire, par convention, à une distance locale inférieure ou égale à 5.

Algorithme 1 : ADAPTO

Entrées : problème cible c , ensembles de règles applicables S_x concluant sur $x = 0, 1, 2$
Sorties : solution du problème, sol_c
Données : $sol_c \leftarrow -1$
si $S_0 \neq \emptyset$ **alors**
 | $sol_c \leftarrow 0$
sinon si $S_1 \neq \emptyset$ **alors**
 | $sol_c \leftarrow 1$
sinon si $S_2 \neq \emptyset$ **alors**
 | $sol_c \leftarrow 2$
fin

Algorithme 2 : ADAPT12

Entrées : problème cible c , ensembles de règles applicables S_x concluant sur $x = 0, 1, 2$
Sorties : solution du problème, sol_c
Données : $sol_c \leftarrow -1$
si $S_0 \neq \emptyset$ **alors**
 | **si** $S_2 \neq \emptyset$ **alors**
 | | $sol_c \leftarrow 2$
 | **si** $S_2 = \emptyset$ **alors**
 | | $sol_c \leftarrow 0$
sinon si $S_1 \neq \emptyset$ **alors**
 | $sol_c \leftarrow 1$
sinon si $S_2 \neq \emptyset$ **alors**
 | $sol_c \leftarrow 2$
fin

La substitution se fait après une étape de sélection des différentes solutions proposées par les règles pouvant être appliquées à un même problème cible. Cette sélection

Algorithme 3 : ADAPT3

Entrées : problème cible c , ensembles de règles applicables S_x concluant sur $x = 0, 1, 2$

Sorties : solution du problème, sol_c

Données : $sol_c \leftarrow -1, n_0 = n_1 = n_2 = 0$

pour $x = 0$ à $2 \wedge S_x \neq \emptyset$ **faire**

 | $n_x \leftarrow$ nombre moyen de prémisses pour les règles de S_x

fin

si $\exists i, n_i > 0 \wedge n_i > n_j \forall j \neq i$ **alors**

 | $sol_c \leftarrow i$

sinon si $n_2 > 0 \wedge n_2 \geq n_j \forall j \neq 2$ **alors**

 | $sol_c \leftarrow 2$

sinon si $n_1 = n_0 \neq 0$ **alors**

 | $sol_c \leftarrow 1$

fin

repose sur un algorithme correspondant à un contexte d'adaptation comme précisé ci-dessus. En pratique, trois algorithmes d'adaptation ont été implantés dans le système. L'algorithme pessimiste (ADAPT0) sélectionne en priorité les règles dont la conclusion vaut 0 (voir algorithme 1); l'algorithme optimiste (ADAPT12) sélectionne d'abord les règles dont la conclusion vaut 2 s'il existe des règles concluant sur 0, puis celles concluant sur 0, et finalement les règles dont la conclusion vaut 1, sinon (voir algorithme 2); l'algorithme pondéré (ADAPT3) sélectionne les règles applicables au problème cible ayant le plus grand nombre de prémisses (voir algorithme 3). Le choix de l'algorithme est laissé à l'utilisateur selon ses objectifs de prédiction. Dans le cas où aucune règle ne s'applique, l'adaptation ne peut être réalisée et on recopie la solution du cas source remémoré. Si plusieurs cas sources ont été remémorés, l'adaptation est réalisée pour chacun successivement et le résultat est choisi par un vote majoritaire.

Pour exemple, considérons que l'ensemble de règles décrit ci-dessus (**R1** à **R4**) est utilisable pour résoudre un problème cible ayant les attributs suivants :

```
{(taille:moyen);(régime hydrique:humide);
(excès d'eau:inondable);(drainage:inexistant);
(occupation précédente:jachère)}
```

L'ensemble de règles adaptables pour ce problème, sous le seuil de distance 0, est alors **{R1, R2, R4}**. L'algorithme ADAPT0 renvoie la solution 0 (application de la règle **R1**). L'algorithme ADAPT12 renvoie la solution 2 (règle **R2**), tandis que l'algorithme pondéré ADAPT3 sélectionne la règle **R4**, dont les deux prémisses sont vérifiées, et renvoie donc la solution 1. Sous le seuil de distance 5, la règle **R3** est adaptable, en effet les valeurs *inondable* et *résurgences et sources* sont à une distance de 5 pour l'attribut *excès d'eau*. Cette règle a le même nombre de prémisses que la règle **R4**, mais cela ne modifie pas le résultat de l'algorithme ADAPT3 qui, en cas d'égalité, favorise les solutions 2 puis 1.

5. Implantation et tests

Le prototype SAMM a été implanté en java en utilisant l’environnement de développement Eclipse³. Dans l’état actuel, l’utilisateur peut charger différentes bases de connaissances et bases de cas, permettant notamment d’appliquer le modèle sur différents territoires, sous réserve de certaines précautions d’usage. L’utilisateur peut aussi facilement tester différents algorithmes et paramètres de remémoration et d’adaptation grâce à un volet de configuration. L’utilisateur peut également évaluer les solutions construites *via* l’interface du système ou à partir des fichiers produits. La figure 4 situe les différents algorithmes implantés et les différentes bases utilisées par SAMM dans le cycle général du RÀPC (Aamodt, Plaza, 1994).

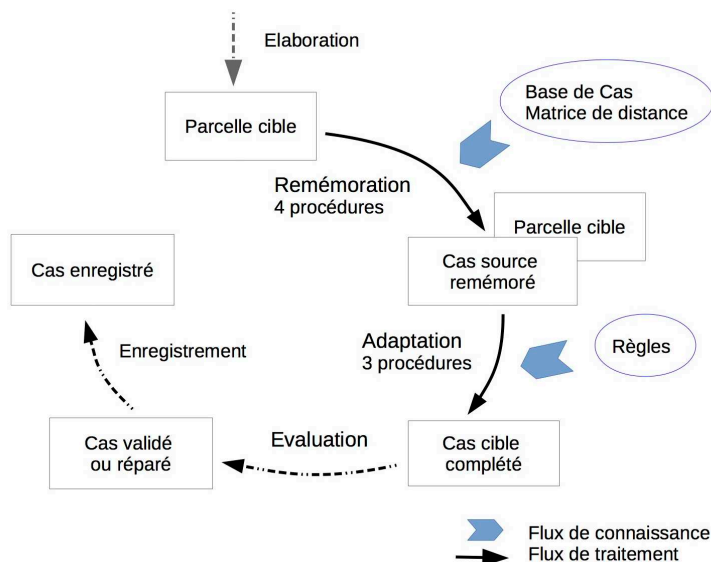


Figure 4. Cycle du RÀPC dans SAMM – en pointillé les étapes non automatisées (Le Ber et al., 2017)

Le fonctionnement du système a ensuite été testé. Pour cela nous avons constitué une sous-base de 72 cas (18 à solution 0 – pas d’implantation, 28 à solution 1 – implantation, et 26 à solution 2 – miscanthus implantable sous conditions). Les tests ont été élaborés puis mis en œuvre à partir de différents paramètres d’expérimentation :

- le choix de l’algorithme de remémoration, décliné en quatre modalités correspondant aux combinaisons des mesures et procédures définies ci-dessus ;
- le choix de l’ensemble des règles utilisables, avec deux modalités : les règles de l’agriculteur du cas source remémoré (ensemble RAS pour *règles agriculteur source*)

3. www.eclipse.org

et les règles des agriculteurs de tous les cas sources de la base (ensemble RAT *règles agriculteur tous*);

- le seuil de distance (fixé à 0, SELECT0, dans cette expérimentation) entre prémisses et descripteurs du problème cible pour le choix de la règle à appliquer;
- le choix de l'algorithme d'adaptation, décliné en trois modalités, correspondant aux trois algorithmes présentés ci-dessus.

Les tests ont été menés en extrayant de la base un individu qui constitue le problème cible, alors que le reste de la base constitue les cas sources à remémorer. La manipulation est répétée 72 fois pour chaque cas de la base.

Dans un premier temps on examine les résultats à l'issue de la remémoration, par simple recopie de la solution (obtenue par vote majoritaire) des cas sources remémorés. La solution proposée est ensuite comparée à la solution originelle du cas cible (le potentiel d'implantation de la parcelle). Les taux de précision et de rappel sont calculés pour chaque procédure et chaque solution (0, 1, ou 2). Les taux de précision sont plus élevés que les taux de rappel, en partie du fait que certains problèmes n'ont pas de solution affectée, quand il n'y a pas de majorité (cela représente 2 à 8 % des cas selon les procédures). Par la suite on présente le taux de rappel moyen pour une procédure donnée.

Sur la figure 5, ces résultats sont présentés en considérant une validation stricte ou relâchée du système. Une validation stricte consiste à classer comme « vraie » toute solution proposée égale à la solution observée 0, 1 ou 2. Une validation relâchée repose sur le regroupement des solutions 1 et 2 : une solution proposée 1 (resp. 2) est classée comme vraie pour une solution observée 2 (resp. 1). On observe que, à l'issue de la remémoration, 52 à 57 % des solutions proposées par le système sont correctes pour la validation stricte, et 61 à 69 % pour la validation relâchée. Les trois algorithmes RM1, RM3, RM4 ont des résultats proches pour la validation stricte ; on remarque que ces algorithmes appartiennent à une moitié ou plus de problèmes cibles avec plusieurs cas sources, ce qui peut conforter la solution proposée, du fait du vote majoritaire. Pour la validation relâchée, le meilleur taux de rappel est obtenu par l'algorithme RM1 qui optimise d'abord la distance globale $d_g(s, c)$ sur l'ensemble DP des descripteurs du problème, puis le nombre d'attributs partagés $n_{DP}(s, c)$. Ces taux sont satisfaisants dans une première étape et compte tenu de la diversité des attributs-valeurs considérés.

Dans un deuxième temps on examine l'effet des algorithmes d'adaptation en jouant sur les différents paramètres cités ci-dessus. Nous avons construit 24 procédures combinant ces paramètres. On s'intéresse ici non pas à la précision et au rappel mais au comportement des algorithmes d'adaptation en fonction de l'algorithme de remémoration (RM1 à 4) et de l'ensemble de règles choisi (RAS ou RAT). Pour cela nous mesurons deux paramètres, le taux d'adaptation, c'est-à-dire le nombre de cas où la solution a effectivement été adaptée, et la part de chaque solution (0, 1, 2 ou indéterminée) attribués aux cas adaptés.

Les taux d'adaptation varient de 40 à 50 % quand on utilise les règles de l'agriculteur source (RAS) et montent à 93 % quand on utilise les règles de tous les agriculteurs

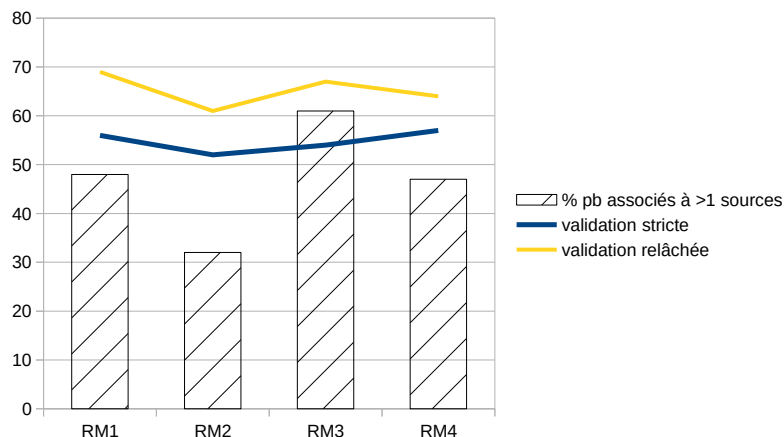


Figure 5. Comparaison des quatre algorithmes de remémoration : taux de rappel moyen et pourcentage des problèmes appariés à plusieurs cas sources

(RAT). Logiquement, ce dernier ensemble couvre à peu près toutes les configurations parcellaires et il y a donc presque toujours une règle qui peut s'appliquer (à l'exception de quelques cas particulier). On est alors proche du fonctionnement d'un système à base de règles. En revanche quand on considère seulement les règles de l'agriculteur source (ou des agriculteurs sources en cas d'appariement multiple), dans plus d'un cas sur deux, il n'y a pas de règle applicable. Le meilleur taux 50 % est obtenu pour l'algorithme RM3, qui optimise d'abord la distance globale $d_g(s, c)$ sur l'ensemble DI (descripteurs de l'index), puis le nombre d'attributs partagés $n_{DI}(s, c)$. Ceci est à relier au fait que les paramètres de l'index d'un cas proviennent des prémisses des règles de l'agriculteur associé à ce cas. À l'inverse, le plus bas taux est obtenu pour RM2 qui optimise d'abord le nombre d'attributs partagés $n_{DP}(s, c)$ (descripteurs du problème).

Les répartitions des différentes solutions pour ces procédures sont présentées dans la figure 6 (seuls sont présentés les résultats obtenus avec le paramètre RAS, le paramètre RAT renvoyant des résultats similaires pour tous les algorithmes de remémoration). Les résultats montrent que l'algorithme pessimiste ADAPT0 permet de favoriser la solution 0 surtout pour RM1 et RM3 (avec des taux d'affectation de 55-56 %). L'algorithme optimiste ADAPT12 favorise la solution 2 pour RM1, RM2 et RM3 (avec des taux d'affectation de 47 à 52 %) et moins pour RM4. Enfin l'algorithme pondéré ADAPT3, qui favorise aussi la solution 2, a un comportement assez proche de ADAPT12, avec un taux de solution 1 qui peut être plus élevé (26 contre 19 % pour RM2). Pour toutes ces procédures, la solution 1 est rarement choisie (de 8 à 26 %) – en partie du fait de la structure des algorithmes (ADAPT0 et ADAPT12 en particulier) et aussi du fait que les règles concluant sur 1 sont très minoritaires dans la base de règles (cf. section 4.2). On remarque également que les solutions indéterminées sont peu nombreuses, ce qui montre que même dans la situation où plusieurs cas sources

sont remémorés, et donc plusieurs jeux de règles sont possiblement utilisés, chaque algorithme d'adaptation renvoie un résultat majoritaire.

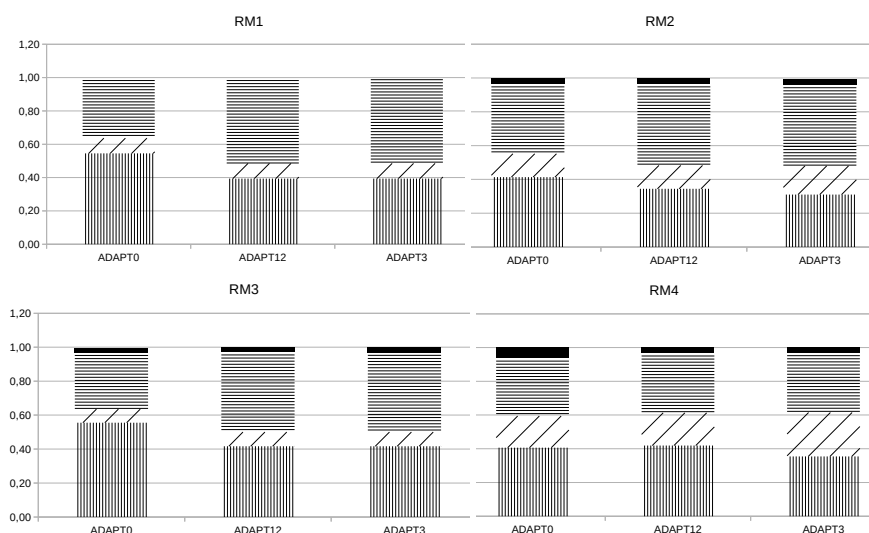


Figure 6. Adaptation : répartition des solutions proposées par les trois algorithmes, pour chaque procédure de remémoration et en utilisant les règles de l'agriculteur source (en tireté vertical, solution 0, oblique, solution 1, horizontale solution 2 et en noir, solution non déterminée)

D'autres tests sont détaillés dans (Martin, 2014). Ils permettent d'explorer les différents comportements des algorithmes proposés. Les quelques résultats présentés ici mettent en évidence que, en fonction de la stratégie de l'utilisateur ou de l'utilisatrice, différentes procédures (ou combinaison de paramètres) peuvent être choisies : s'il ou elle cherche à estimer le taux d'implantation du miscanthus dans un contexte négatif (ou minimal), alors il ou elle choisira la procédure RM1-ADAPTO-RAS ou RM3-ADAPTO-RAS ; dans le cas contraire, il ou elle choisira la procédure ADAPT12-RAS avec l'un des trois algorithmes RM1, 2 ou 3. Enfin pour être plus proche des choix potentiels des agriculteurs, il vaut sans doute mieux utiliser la procédure RM4-ADAPT3-RAS. L'utilisation de l'ensemble des règles RAT permet d'amplifier le comportement des algorithmes ADAPTO et ADAPT12.

Finalement cette expérimentation montre le rôle central qui doit être laissé à l'utilisateur ou utilisatrice, pour choisir les algorithmes d'une part, mais aussi pour examiner les résultats obtenus aux différentes étapes : cas remémorés, règles applicables, traitement des solutions indéterminées. Nous montrons aussi que l'usage du prototype SAMM est plus large et varié que la simple prédiction, la plus juste possible, d'une localisation du miscanthus. C'est ce que nous allons discuter dans la suite.

6. Exemples et discussion

Pour expliquer les comportements du système révélés dans les tests présentés ci-dessus, nous détaillons dans cette partie quelques cas similaires selon leurs index et qui ont cependant des solutions différentes, c'est-à-dire que les parcelles ont des potentiels d'implantation du miscanthus différents. Ces exemples soulignent l'intérêt d'explorer davantage les critères de choix et les règles de décision des agriculteurs, qui s'expriment à différentes échelles spatiales. Nous discutons ensuite de l'intérêt de l'approche proposée dans le système SAMM et de la manière de rendre ce système plus exploitable, dans le sens d'un système de raisonnement à partir d'explications.

6.1. Des voisinages arborés

Dans la base se trouvent les trois cas suivants, dont les attributs index et la solution sont donnés ci-dessous. L'ensemble des attributs de chaque cas est décrit au tableau 3.

Cas 1a Numéro = 39847961

Index = {(forme : irréguliere);
(voisinage arbres : bois)}
Solution = 1

Cas 2a Numéro = 10533211

Index = {(forme : étroite);
(voisinage arbres : bois)}
Solution = 0

Cas 3a Numéro = 10915621

Index = {(forme : contraignante);
(taille : petit); (voisinage arbres : forêt)}
(régime hydrique : partiellement hydromorphe);
Solution = 1

Ces trois cas admettent tous l'attribut *voisinage arbres* dans l'index, avec des valeurs proches (*bois* ou *forêt*), mais pour autant ils n'admettent pas tous la même solution. Si on examine les éléments d'explication donnés par les agriculteurs on peut dire : dans le cas **1a**, l'implantation de miscanthus sur une parcelle *biscornue* en lisière de forêt est possible ; dans le cas **2a**, la parcelle est étroite, ce qui génère trop d'ombre sur la parcelle et donc pas assez de surface avec un réel potentiel de production du miscanthus : l'agriculteur décide de ne pas planter ; dans le cas **3a**, pour une parcelle de petite taille et de forme contraignante, l'agriculteur préfère mettre du miscanthus et assumer le risque de faibles rendements plutôt que de continuer à travailler une parcelle très peu rentable pour la rotation classique. Dans ce dernier cas, l'agriculteur utilise davantage d'éléments pour expliquer sa décision (4 descripteurs pour l'index).

Tableau 3. Description de trois cas ayant l'attribut *voisinage arbres* dans leur index – en italique les attributs retenus dans les index

	Cas 1a	Cas 2a	Cas 3a
Attribut	valeur	valeur	valeur
Occupation préc.	rotation culturale		rotation culturale
Potentiel agro.	bon		bon
Mécanique du sol	partiellement battant		
Forme	<i>irrégulière</i>	<i>étroite</i>	<i>contraignante</i>
Voisinage arbres	<i>bois</i>	<i>bois</i>	<i>forêt</i>
Voisinage hydro.	cours d'eau	cours d'eau	
Régime hydrique		humide	<i>partiel. hydromorphe</i>
Protection env.		natura 2000	
Taille			<i>petit</i>

6.2. De la distance au siège

Nous examinons maintenant deux cas qui possèdent des caractéristiques communes ou proches mais qui néanmoins sont traités de manières différentes par les agriculteurs. Ici c'est l'organisation du parcellaire dans l'exploitation qui va principalement jouer sur la décision. L'ensemble des attributs de chaque cas est décrit au tableau 4.

Cas 1b Numéro = 10437971

```
Index = {(forme:pointue ET convexe ET étroite)
(taille:très petit)
(distance au siège:loin)
(distance autre miscanthus:loin)
(accès largeur:étroit)
(largeur franchissement :très étroit)
(voisinage habité:oui) }
Solution = 0
```

Cas 2b Numéro = 10917151

```
Index = {(forme:contraignante);
(taille:très petit);
(distance au siège:loin);
(carrossabilité:peu carrossable);
(accès largeur:étroit);
(voisinage arbres:foret);
(régime hydrique:hydromorphe)}
Solution = 1
```

Dans les deux cas, il s'agit d'une parcelle difficile d'accès et à forme et taille contraignantes. La parcelle est implantée en miscanthus pour le cas **2b** mais pas pour le cas **1b** où l'agriculteur souhaite optimiser ses chantiers de récolte du miscanthus : il

ne veut pas implanter du miscanthus sur des parcelles trop éloignées entre elles comme l'indique l'attribut *distance autre miscanthus* qui a pour valeur *loin*. On voit ici que la prise de décision s'appuie sur les caractéristiques de la parcelle mais aussi sur l'organisation du territoire de l'exploitation agricole.

Tableau 4. Description de deux cas – en italique les attributs retenus dans les index

	Cas 1b	Cas 2b
Attribut	valeur	valeur
Occupation préc.	jachère	rotation culturale
Dist. siège	<i>loin</i>	<i>loin</i>
Forme	<i>pointue</i> <i>ET convexe</i> <i>ET étroite</i>	<i>contraignante</i>
Taille	<i>très petit</i>	<i>très petit</i>
Dist. Miscanthus	<i>loin</i>	
Accès largeur	<i>étroit</i>	
Carrossabilité		<i>peu carrossable</i>
Accès franchissement	entre maisons	
Largeur franchissement	<i>très étroit</i>	
Régime hydrique		<i>hydromorphe</i>
Voisinage hab.	<i>habité</i>	
Voisinage arbres		<i>forêt</i>

6.3. Quel modèle de RÀPC pour les cas d'allocation du miscanthus ?

Le raisonnement à partir de cas est essentiellement connu pour le raisonnement analogique direct, consistant à s'inspirer d'un problème résolu (le cas source) pour l'adapter afin de résoudre un nouveau problème (le cas cible). Une des bonnes propriétés du RÀPC est qu'il n'est pas nécessaire d'avoir de très nombreux cas, mais plutôt des cas couvrant bien l'espace des solutions, ce qui requière aussi une certaine indépendance des attributs vis-à-vis de la décision et un domaine homogène. En complément, la condition nécessaire et suffisante pour raisonner à partir d'un cas est de disposer des connaissances d'adaptation, elles-mêmes impliquant les connaissances de similarité (Fuchs *et al.*, 2014).

Dans l'application qui nous occupe ici, les premières observations sur le terrain ont montré qu'il n'était pas possible de découvrir directement ces connaissances d'adaptation, ce qui a conduit à reformuler les questions posées aux agriculteurs, dans le but de comprendre les raisons qui font qu'un exploitant choisit ou non d'implanter la nouvelle culture (le miscanthus) dans une parcelle de son exploitation. Plutôt que le résultat du choix, les *explications* – que nous avons utilisées dans SAMM pour construire les règles servant à l'adaptation – de ce choix positif ou négatif pourraient constituer alors la partie solution d'un cas. Or on a montré dans les exemples ci-dessus que les seuls attributs contenus dans l'index ne suffisaient pas pour expliquer complètement

ces choix, donc pour construire une explication. Il faut encore affiner des relations de dépendance entre la description de la parcelle, celle de l'exploitation et les énoncés explicatifs, pour pouvoir déterminer les différents éléments qui peuvent expliquer un choix positif ou négatif.

Ce type de raisonnement – produisant des explications – a été particulièrement développé dans la recherche d'argumentaire pour justifier la référence à une jurisprudence dans la législation américaine : il s'agit d'adapter les explications d'une jurisprudence gagnante à un cas en cours de jugement (Schank *et al.*, 1994). Dans le cadre de notre étude, ceci revient à reformuler la question de recherche : quels éléments peuvent être donnés à un responsable de coopérative (l'équivalent du juge dans notre cas) pour que, dans la situation qu'il évalue, il anticipe au mieux le choix de l'agriculteur d'adopter ou non la culture nouvelle ? Ce jugement du responsable ne repose pas sur une expertise forte du domaine – puisqu'il s'agit justement d'une culture nouvelle – et il a donc besoin de s'appuyer sur les raisons expliquant les choix d'autres agriculteurs dans la même situation. On pourrait alors envisager une approche telle que celle développée dans (Le Ber *et al.*, 2011) pour capitaliser l'expérience de porteurs d'observatoires de pratiques agricoles.

Le raisonnement à partir de cas pourrait être exploité pour une autre excellente propriété qu'il exhibe : le RÀPC est non seulement une méthode de résolution de problèmes, mais aussi une méthode d'apprentissage incrémental de connaissances. Dans le cycle du RÀPC, l'apprentissage se fait au moment du bilan d'un cas, de sa réparation et d'une nouvelle tentative avec la mobilisation de nouvelles connaissances. Les connaissances principalement découvertes sont les connaissances de dépendance entre descripteurs du problème et descripteurs de la solution à ce problème. Ces connaissances sont nécessaires pour construire les connaissances d'adaptation et de similarité. Cette boucle d'apprentissage est ouverte aux experts rassemblés à cette occasion pour mobiliser leurs connaissances et corriger, et compléter les connaissances d'adaptation. Il est alors intéressant d'utiliser le RÀPC comme médium de construction collaborative de connaissances à partir des réparations de cas précis. Au-delà de la constitution d'une base de cas sans doute bien plus facile à réutiliser au fur et à mesure des apprentissages, les connaissances des experts (ici par exemple agriculteurs et responsables de coopérative travaillant ensemble) échangent et contribuent à l'intégration mutuelle de connaissances jusqu'ici non partagées.

7. Conclusion et perspectives

À ce stade et d'un point de vue thématique, pour les agronomes, le développement de SAMM a permis deux avancées majeures : (i) formaliser les règles de décision des agriculteurs à partir des résultats d'enquêtes complexes ; (ii) construire, en vue de prospective, des gammes de scénarios d'allocation du miscanthus. Toutefois, une difficulté demeure : les cas sont décrits par un petit nombre d'attributs, ce qui rend les appariements fragiles (les problèmes cible et source ne partagent le plus souvent que 2 ou 3 descripteurs). Ceci renforce la nécessité de bien choisir les règles qui per-

mettent l'adaptation ou tout au moins de les choisir en connaissance de cause. Par exemple pour la prospective, on peut envisager un scénario conduisant à une faible production (les exploitants n'implantent du miscanthus que dans certaines parcelles, les plus adaptées) – on utilise alors l'algorithme pessimiste ADAPT0 – ou un scénario conduisant à une forte production (les exploitants implantent dès qu'ils le peuvent) – on utilise l'algorithme optimiste ADAPT12. D'autre part, les exemples que nous avons examinés ci-dessus (section 6) montrent que le raisonnement des agriculteurs s'effectue à plusieurs échelles (la parcelle, le groupe de parcelle, l'exploitation). Il faudrait affiner en ce sens la description et la sélection des règles afin d'avoir de meilleurs résultats. Ces résultats pourraient alors se présenter comme différentes options : en ne considérant que la parcelle, ou son proche voisinage ou l'exploitation dans son ensemble par exemple.

De fait, on se trouve ici dans une situation d'utilisation de l'expérience qui relève d'un raisonnement à partir d'explications. En effet, s'il semble possible d'élucider quelques explications générales sur les décisions (les règles énoncées par les agriculteurs pour expliquer leur décision), les cas eux-mêmes sont difficiles à adapter et les agriculteurs ne se sont pas prêtés au jeu d'utiliser un cas « similaire » pour dire ce qui est différent entre leur cas et un autre cas pouvant expliquer pourquoi la décision est différente ou la même. La complexité du processus de raisonnement et l'importance de donner la main aux décideurs plaident dans le sens d'explorer ces différences avec les décideurs. Ce qui ressort vraiment clairement du travail réalisé, c'est que la description du cas et de ses explications est une excellente façon de rendre concrètes les connaissances issues des enquêtes. L'utilisation du raisonnement est intéressante spécifiquement pour les interactions qu'il permet, plutôt que pour sa capacité de prédiction. Le prototype est utile pour construire de nouvelles boucles d'enquête et de réalisation de prototypes qui seront cette fois vraiment orientés sur la prédiction ou l'aide à la décision.

La notion d'explication semble donc la plus intéressante à considérer dans l'exploration de connaissances décisionnelles et, en particulier, dans ce contexte évolutif de l'introduction dans les territoires d'exploitations agricoles d'une nouvelle culture aux caractéristiques et enjeux spécifiques. L'association directe de la résolution et de l'apprentissage pour amorcer une base de cas et construire les conditions de son efficacité (appropriation du processus par les acteurs concernés) constitue une perspective intéressante pour le projet présenté dans cet article, tout en permettant la formation mutuelle des experts et des exploitants s'appropriant les connaissances des uns et des autres.

Dans cette perspective, on pourrait alors aussi proposer un système d'assistance à l'agriculteur, pour l'adaptation de son exploitation à l'implantation du miscanthus, avec une solution plus complexe que le simple choix d'adopter ou non la culture (choix de parcelles, réorganisation du travail, ...).

Remerciements

Cet article s'appuie sur le travail de thèse de Laura Martin, financé par le projet FUTUROL et par OSEO. Les enquêtes ont pu être menées grâce à de nombreux agriculteurs, ainsi qu'à M. Béjot (Bourgogne Pellets), M. Garnaud (SCA Déshydratation de la Haute-Seine) et Mmes Pierson et Petit (Chambre d'Agriculture de Bourgogne). Ce travail a également bénéficié des discussions menées dans le cadre du réseau INRA Payote. Enfin, nous remercions les lecteurs de la revue dont les remarques nous ont permis d'améliorer grandement ce texte.

Bibliographie

- Aamodt A., Plaza E. (1994, March). Case-Based reasoning: Foundational Issues, Methodological Variations, and System Approachs. *AICOM*, vol. 7, n° 1, p. 39–59.
- Adeyanju I., Wiratunga N., Lothian R., Sripada S., Lamontagne L. (2009). Case Retrieval Reuse Net (CR2N): An Architecture for Reuse of Textual Solutions. In L. McGinty, D. C. Wilson (Eds.), *Case-based reasoning research and development*, vol. LNCS 5650, p. 14-28. Springer.
- Avesani P., Perini A., Ricci F. (2000). Interactive case-based planning for forest fire management. *Applied Intelligence*, vol. 13, n° 1, p. 41-57.
- Bain W. M. (1989). Judge. In C. Riesbeck, R. Schank (Eds.), *Inside case-based reasoning*, p. 93–140. Lawrence Erlbaum Associates.
- Benoît M., Rizzo D., Marraccini E., Moonen A.-C., Galli M., Lardon S. *et al.* (2012). Landscape agronomy: a new field for addressing agricultural landscape dynamics. *Landscape Ecology*, vol. 27, n° 10.
- Bocquého G. (2012). *Risque, temps et adoption des cultures pérennes énergétiques: exemple du cas français*. Thèse de AgroParisTech.
- Bosch O. J. H., Gibson R. S., Kellner K., Allen W. J. (1997). Using case-based reasoning methodology to maximise the use of knowledge to solve specific rangeland problems. *Journal of Arid Environments*, vol. 35, p. 549–557.
- Brüninghaus S., Ashley K. D. (2001). The Role of Information Extraction for Textual CBR. In D. W. Aha, I. Watson (Eds.), *Case-based reasoning research and development*, vol. LNCS 2080, p. 74-89. Springer.
- Burke R., Kass A. (1996). Retrieving stories for case-based teaching. In D. Leake (Ed.), *Case-based reasoning. experiences, lessons and future directions*, p. 93–109. AAAI Press/The MIT Press.
- Du Y., Liang F., Sun Y. (2012). Integrating spatial relations into case-based reasoning to solve geographic problems. *Knowledge-Based Systems*, vol. 33, p. 111–123.
- Dufour-Lussier V., Le Ber F., Lieber J., Nauer E. (2014). Automatic case acquisition from texts for process-oriented case-based reasoning. *Information Systems*, vol. 40, p. 153–167.
- Fuchs B., Lieber J., Mille A., Napoli A. (2006, mars). Une première formalisation de la phase d'élaboration du raisonnement à partir de cas. In *14ème séminaire sur le raisonnement à partir de cas*.

- Fuchs B., Lieber J., Mille A., Napoli A. (2014, juin). Differential Adaptation: an Operational Approach to Adaptation for Solving Numerical Problems with CBR. *Knowledge Based Systems*, n° 68, p. 103-114.
- Girard N., Bellon S., Hubert B., Lardon S., Moulin C.-H., Osty P.-L. (2001). Categorising combination of farmers' land use practices: an approach based on examples of sheep farms in the south of France. *Agronomie*, vol. 21, p. 435-459.
- Guyet T., Le Ber F., Teisseire M. (2013). Intelligence artificielle et agronomie – introduction au numéro spécial. *Revue d'Intelligence Artificielle*, vol. 27, p. 403-407.
- Hammond K. J. (1990). Case-Based Planning: A Framework for Planning from Experience. *Cognitive Science*, vol. 14, n° 3, p. 385-443.
- Hastings J. D., Latchininsky A. V., Schell S. P. (2010). CARMA: Scalability with Approximate-Model-Based Adaptation. In *2010 International Congress on Environmental Modelling and Software, Ottawa, Canada*, p. 1-10.
- Herbeaux O., Mille A. (1999). Accelere: a case-based design assistant for closed cell rubber industry. *Knowledge-Based Systems*, vol. 12, n° 5-6, p. 231-238.
- Holt A., Benwell G. (1996). Case-Based Reasoning and Spatial Analysis. *Journal of the Urban and Regional Information Systems Association*, vol. 8, p. 27-36.
- Leake D., McSherry D. (2005). Introduction to the Special Issue on Explanation in Case-Based Reasoning. *Artificial Intelligence Review*, vol. 24, n° 2, p. 103-108.
- Le Ber F., Brassac C. (2008). Étude longitudinale d'une procédure de modélisation de connaissances en matière de gestion du territoire agricole. *Revue d'Anthropologie des Connaissances*, vol. 2, n° 2, p. 151-168.
- Le Ber F., Dolques X., Martin L., Mille A., Benoît M. (2017, juin). A Reasoning Model Based on Perennial Crop Allocation Cases and Rules. In *Int. Conference on Case Based Reasoning, ICCBR 2017*, vol. LNAI 10339, p. 61-75. Trondheim, Norway, Springer.
- Le Ber F., Martin L., Mille A., Benoît M. (2015, juillet). Un système de RÀPC pour prévoir l'allocation de parcelles agricoles. In D. Lourdeaux (Ed.), *Ire conférence sur les Applications Pratiques de l'Intelligence Artificielle (APIA)*. Rennes, France.
- Le Ber F., Napoli A., Metzger J.-L., Lardon S. (2003). Modeling and comparing farm maps using graphs and case-based reasoning. *Journal of Universal Computer Science*, vol. 9, n° 9, p. 1073-1095.
- Le Ber F., Nogry S., Brassac C., Benoît M. (2011). Capitalisation d'expériences pour la mise en place d'observatoires de pratiques agricoles — une approche fondée sur l'exploitation de narrations. *Revue Internationale de Géomatique*, vol. 21, n° 1, p. 99-118.
- Martin L. (2014). *Modéliser l'insertion territoriale du Miscanthus x giganteus à partir des décisions des agriculteurs : une approche exploitant le modèle du raisonnement à partir de cas*. Thèse de l'Université de Lorraine.
- Martin L., Le Ber F. (2013). Raisonner à partir de cas (ou pas) pour l'allocation de cultures pérennes dans un territoire agricole. In *21e atelier Français de Raisonnement à Partir de Cas, Lille*, p. 1-9.
- Martin L., Wohlfahrt J., Le Ber F., Benoît M. (2012). L'insertion territoriale des cultures biomasses pérennes. Etude de cas sur le miscanthus en Côte d'Or (Bourgogne, France). *L'Espace Géographique*, n° 2, p. 138-153.

- Osty P.-L., Le Ber F., Lieber J. (2008). Raisonnement à partir de cas et agronomie des territoires – constructions croisées. *Revue d'Anthropologie des Connaissances*, vol. 2, n° 2, p. 169–193.
- Riesbeck C. K., Schank R. C. (1989). *Inside Case-Based Reasoning*. Hillsdale, New Jersey, Lawrence Erlbaum Associates.
- Rizzo D., Martin L., Wohlfahrt J. (2014). Miscanthus spatial location as seen by farmers: A machine learning approach to model real criteria. *Biomass and Bioenergy*, vol. 66, p. 348–363.
- Roth-Berghofer T. (2004). Explanations and case-based reasoning: Foundational issues. In *Advances in Case-Based Reasoning*, p. 389-403. Springer.
- Schank R. C., Kass A., Riesbeck C. K. (Eds.). (1994). *Inside case-based explanation*. Hillsdale, New Jersey, LEA.
- Schnitzler A. (2011). *Miscanthus: L'homme cultive-t-il un nouvel envahisseur?* Rapport, Université Paul Verlaine de Metz, Laboratoire des Interactions Ecotoxicologie, Biodiversité, Ecosystèmes.
- Sørmo F., Cassens J., Aamodt A. (2005). Explanation in Case-Based Reasoning – Perspectives and Goals. *Artificial Intelligence Review*, vol. 24, n° 2, p. 109-143.

