A Reasoning Model Based on Perennial Crop Allocation Cases and Rules

Florence Le Ber, Xavier Dolques, Laura Martin, Alain Mille, Marc Benoît

A Reasoning Model based on Perennial Crop Allocation Cases and Rules

Florence Le Ber[1], Xavier Dolques[1], Laura Martin[2], Alain Mille[3], and
Marc Benoît[2]

(1) ICube, Université de Strasbourg, ENGEES, CNRS, France
florence.leber,xavier.dolques@engees.unistra.fr
(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

Abstract. This paper presents a prototype of case-based reasoning,
built for the agricultural domain. Its aim is to forecast the allocation
of a new energy crop, the miscanthus. Interviews were conducted with
french farmers in order to know how they make their decisions. Based on
interview analysis, a case base and a rule base have been formalized, to-
gether with similarity and adaptation knowledge. Furthermore we have
introduced variations in the reasoning modules, for allowing different
uses. Tests have been conducted. Results showed that the model can be
used in different ways, according to the aim of the user, and e.g. the
economic conditions for miscanthus allocation.

Keywords: Case Based Reasoning, Adaptation, Decision rules, Agri-
culture

1 Introduction

To face the decrease of fossil energy supplies and to reduce the greenhouse gas
emissions, new biomass energy[1] resources become of a great interest in Europe,
Their spatial extension seems then unavoidable. For instance, the miscanthus
(*Miscanthus ex giganteus*), has interesting caloric values and constitutes a great
potential for biofuel and heating plant. As the production of such biomass crops
is perennial (15 to 20 years of production by land) and exclusively dedicated
to the energy use, it is necessary to anticipate their allocation to prevent a
forecasted perennial food / non-food competition.

Modeling is useful to anticipate the extension of biomass crops and bring
decision-making support for politics. Several land-use change models deal with
this specific problem [1]. Most of these models simulate large-scale allocation pro-
cesses, taking into account numerous biophysical variables but only few human

---

[1] Biomass energy corresponds to organic matter, essentially from agricultural and
forest products (e.g. sugar beet, wood), co-product (e.g. wheat straw) and wastes
(e.g. liquid manure).

decision-making processes linked to the land system management of farmers, whereas it is a major driving factor of miscanthus allocation process [2]. Indeed, representing and modelling human behaviour and decisional processes regarding land use change is difficult and constitutes a main research challenge [3].

The goal of our research is thus to model the miscanthus allocation according to farmers practices and decision-making process. Because the allocation of miscanthus is too recent to use national statistics, we decided to build our model based on a case study in Burgundy (East of France). Furthermore, because the allocation of miscanthus is too recent to be fully understood, we decided to rely on a case-based reasoning approach [4, 5] as a pathway to use current practices for predicting land use change.

Case-based reasoning (CBR), introduced in [6], indeed allows to model application domains where general knowledge is incompletely formalized and where expertise mainly relies on experiences. Examples of such domains are medicine, chemistry, engineering, risks prevention or cooking, domains where case based reasoning systems have been successfully implemented.

Building a CBR system requires primarily to work on knowledge modeling. This step can rely on documents, that are more or less normalized, like a patient file, or a cooking recipe (examples are given in [7]). Data mining or text analysis techniques can also be used [8]. However, in several application domains, a long collaborative analysis work is required between computer scientist and domain experts. In our application, knowledge modeling has first required to survey farmers who had -or not- chosen to plant Micanthus in their farm fields. These surveys have resulted in two types of knowledge: on one hand, knowledge on farm field characteristics, as described by farmers, on the other hand, explanations given by farmers on the reasons why they choose or not to plant miscanthus in a specific field, and on conditions that could change their decision [2].

Relying on these pieces of knowledge, we have built a CBR proptotype, named SAMM, i.e. Spatial Allocation Modelling of Miscanthus, which aims at forecasting miscanthus allocation in farm fields. This prototype is composed with a case base, a rule base, a knowledge base (containing both similarity and adaptation knowledge) and a reasoning module, including various strategies.

The paper is structured as follows. Section 2 presents an overview of related work. The SAMM prototype is detailed in Sections 3 and 4. Experiments are conducted in Section 5. The paper ends with a conclusion and some perspectives.

2 Related work

Our work belongs to a research domain at the intersection of artificial intelligence and agronomy. More precisely, it relies on a previous work [9] which focused on the comparison of farm surveys and proved the interest of CBR as a modeling tool for landscape agronomy [10]. In this case model, the problem was a farm spatial organization (e.g., location of farm plots, roads, farm buildings) and the solution was a farm functional organization. The assumption was that similar spatial organizations corresponded to similar functional organizations.

More largely, there exists numerous systems linking CBR and environmental sciences, most of them based on numerical approaches, close to machine learning methods. In [11] for instance, spatial relations between neighboring areas are used to compute a similarity measure between them and forecast their land use (buildings, forest or crops). Older work already used CBR to analyse geographic data, e.g. for soil classification [12].

When CBR systems are based on numerical approaches, knowledge of stakeholders is slightly included, whereas our objective is to mainly rely on this knowledge. The system described in [9] is based on stakeholder knowledge that was already synthesized by researchers. The CARMA system is used for diagnosis and treatment of crop destructive animals [13], by adapting models built on expert knowledge, and was generalised over several american states. [14] handled stakeholder knowledge within a modeling approach close to CBR, but no system was implemented, due to the complexity of forms of knowledge to be modelled (management of sheep herds). There are actually very few systems that explicitely includes stakeholder experiences to be shared, as done in the system described by [15], which gathers community knowledge about rangeland management in New-Zealand.

Our work also deals with explanations and thus can be linked to *Explanation Based Reasoning* [16, 17]. Indeed, in our work, farmer explanations about their choice are modelled by rules, that can be used to propose –what is done by the current prototype– and explain a solution –what could be done in the future. Such a reasoning approach has been initially developed for argumentation building based on american case law [6]. To our knowledge no application to the agricultural domain has been developed sofar.

A machine learning approach has been used to extract statistical rules explaining the spatial location of miscanthus, based on the characteristics of farm fields [18]. On the contrary, we have chosen to developp a CBR approach that rely on decision rules stated by farmers.

3   Case and Knowledge bases

The originality in SAMM is to use both a case and a rule base. Rules are linked to cases and used for the adaptation step.

3.1   The case base

In our application, a case is defined as a specific experience of miscanthus allocation (or non allocation) in a farm field. The problem-solution pair is a farm field and its allocation potential for miscanthus. Each case is represented with a vector of qualitative values, divided in two parts:

1. the problem part gives the farm field characteristics, as described by the farmer; there are 32 possible attributes with 159 values influencing the allocation potential of miscanthus, classified into the six following categories:

- agronomy, 14 attributes, e.g. last land-use, soil characteristics, agronomic potential, slope, soil water regime of the field;
- geometry, 2 attributes, shape and size of the field;
- access, 7 attributes, e.g. distance to farmstead, crossing zones;
- neighbourhood, 6 attributes, e.g., tree, village, crop neighbourhood;
- ownership, 2 attributes, land status and perennial use of the field;
- environmental measures, 1 attribute, protected site.

The subset of attribute-values describing the problem is denoted DP (for *problem descriptors*) and can be formalized as a set of pairs $(a, v) \in \mathcal{A} \times \mathcal{V}$, where $\mathcal{A}$ and $\mathcal{V}$ are respectively the attribute set and the value set.

2. the solution part describes the miscanthus allocation potential of the farm field with a unique variable and three values: the field cannot be allocated (value 0), can be allocated (value 1) or can be allocated under conditions (value 2); then the solution is formalized as a pair (miscanthus, i) where $i = 0, 1$ or 2.

To each case is also associated a label refering to the farmer who manages the farm field.
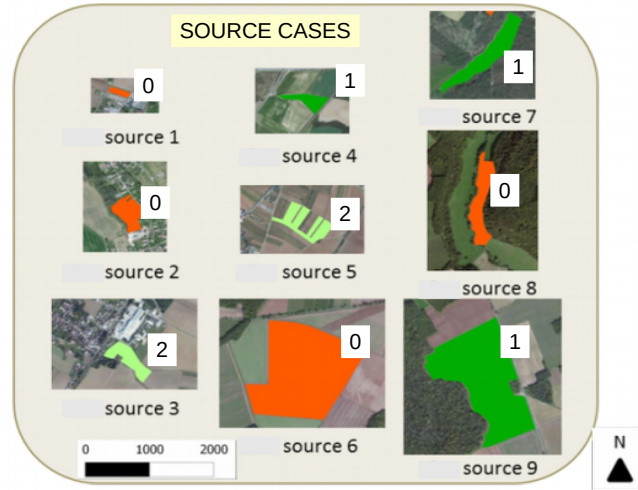
Fig. 1. Some farm fields that are modeled as source cases; the allocation potential is represented with a color (0=red, 1=green, 2=pale green)

The case base in SAMM includes 82 farm fields of which micanthus allocation potential has been stated by farmers during past interviews. For these problems, the solution is known. They are called *source cases* in the following (see examples on Figure 1). The farmer associated to a source case is called *source farmer*. The case base organisation is flat, that is all cases belong to a same information level.

Nevertheless, to help retrieval and adaptation steps, indexes are used. Each case is thus described with a subset of descriptors, denoted by DI (for *index descriptors*), collecting the elements that were explicitly involved in the farmer decision process. For instance, when a farmer says: *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"*[2], attributes about the soil water regime of the farm field and its last land-use are included in the set of index descriptors.

Attributes	Source 1 DP	Source 1 DI	Source 2 DP	Source 2 DI
last land-use	crop rotation		crop rotation	
agronomic potential	good		middle	
soil depth			shallow	
stony soil			high	
soil mechanics	partial crusting			
area morphology			steep-sided	x
field shape	irregular	x	irregular	
tree neighbourhood	wood	x	wood	x
water neighbourhood	watercourse			
land status	property		renting	
allocation potential	1		0	

Table 1. Two source cases – DP: problem descriptors; DI: index descriptors (marked with x); attributes describe different aspects of the fields: agronomy, geometry, neighbourhood, and ownership

Two case examples are described in Tab. 1. Only attributes with values for each case are represented, i.e. attributes which were mentioned by the farmer to describe his/her field. The value of other attributes are not known for these farm fields. Note that there are few problem descriptors (DP) (6 or 7 among 32 possible) and still less index descriptors (DI), that are used to explain the farmer decision (here 2 or 3, marked by an x). This involves that most attributes are sparsely represented in the case base: 14 attributes (among 32) are used in less than 9 source cases, 10 are used in 10 to 19 cases and 8 are used in almost or more than half of the source cases. Finally cases have generally few attributes in common. All attributes are discrete (some are binary) since the case descriptions come from the analysis of farmer interviews, and thus are based on verbal nominal data. To elaborate the problems to be solved (called *target problems*), it is necessary to link these nominal attributes with numerical data, like geographical data. Furthermore, it can be difficult to obtain some informations (e.g. the loca-

---

[2] A not drained field: very bad, that is wet, wet, very wet (...) I always saw fallows there (...) it is not the same price: I said "hop, I make miscanthus inside"

tion of the farmstead) but others are frequently used in agronomical applications (last land-use, farm field geometry, soil characteristics and water regime, etc.).

3.2 The rule base

The SAMM reasoning system also relies on a set of rules that have been collected after transcription of farmer interviews [2]. They formalize the elements given by farmers when explaining their decision to plant (or not) miscanthus in a farm field. These rules are called decision rules. They are of two types: generic rules are independant from space (they are non-spatial) while spatial rules are linked to a farm field. Non-spatial rules are for example about economical context or environmental regulation; spatial rules are concerned with a farm field own characteristics. Rules are represented as pairs $< \bigwedge$ `conditions` (descriptors), `conclusion` (allocation potential) $>$. Conditions depends on various attributes and values as examplified in the two following rules: $<$ (`distance to farmstead` $=$ `close`) $\wedge$ (`access suitability` $=$ `low`) $\wedge$ (`protected site` $=$ `Natura 2000`), $(1) >$ and $<$ (`soil water regime` $=$ `flood area`), $(0) >$.

Each rule is labelled with an identifier refering to the farmer who expressed it, and thus linked to source cases labelled with the same identifier. It is worth noting that the rule set can be inconsistant since farmers can consider the same elements in different ways, i.e. a same field characteristic can have a positive or negative influence on their decision to implant miscanthus. For instance, the rule $<$ (`distance to farmstead` $=$ `far`), $(2) >$ holds for a farmer A_1, whereas another rule $<$ (`distance to farmstead` $=$ `far`), $(0) >$ holds for a farmer A_2. Furthermore, these rules are essential for the agronomists, since they express how and why farmer do choose or not to plant miscanthus.

The rule base covering all surveyed farmers includes 96 rules: 61 rules with conclusion 0 (the farm field cannot be allocated with miscanthus), 8 rules with conclusion 1 (the farm field can be allocated with miscanthus) and 27 rules with conclusion 2 (the farm field can be allocated with respect to further conditions). Regarding the size of rules, 65% of them have only one condition, 25% have 2 conditions and 10% have 3 or more (until 7) conditions.

The low number of rules with conclusion 1 highlights the novelty and scarcity of miscanthus plots. We therefore focus below mainly on the situations where the farm field can be allocated under conditions or cannot be allocated.

3.3 Similarity and adaptation knowledge

Similarity knowledge has been elaborated based on the content analysis of interviews made with farmers. Various levels and types of similarity are considered (in the following s refers to the *source* case and t refers to the *target* case):

- a global numeric level: the number of common descriptors (whatever value) of both *source* and *target* problems, denoted n; the sets DP_s of *source* and DP_t of *target* can be considered, then:

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

or the sets DI_s and DP_t (the set DI is unknown for *target*) are considered, then:

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

– a global semantic level: the global distance between *source* and *target* is the average of local measures (denoted d_l) on common attributes ; it can be computed for $n = n_{DP}$ or $n = n_{DI}$:

$$d_g(s,t) = \Sigma_{i=1}^{n} d_l((a_i, v_{is}), (a_i, v_{it})))/n$$

The local measures on attributes are given in a distance matrix (see an excerpt in Tab. 2). Distances rely on the influences of attribute values towards the miscanthus allocation potential of farm fields (same influence or opposed influence), based on the analysis of farmer interviews. For instance, considering the last land-use, the value "forage crop" is close to the value "crop rotation", but distant from the value "fallows"; finally, distance values have been chosen heuristically as follows:

– $d_l((a, v_s), (a, v_t)) = 5$ if both values have similar influences, positive or negative with respect to miscanthus allocation;
– $d_l = 20$ if both values have different influences, one is positive and the other negative;
– $d_l = 10$ if at least one of the values has a neutral influence.

	forage crop	production	fallows	grassland	crop rotation
forage crop	0	5	20	5	5
production	5	0	20	5	5
fallows	20	20	0	20	20
grassland	5	5	20	0	5
crop rotation	5	5	20	5	0

Table 2. Semantic distance between values of attribute "last land-use"

In CBR systems, the solution adaptation is achieved when at least one descriptor of the source problem is different from the one of the target problem. Transformational adaptation [19] starts from the solution of the selected source case, and modifies it with respect to the differences between the source and the target problems.

To build such a solution, we here rely on the farmer decision rules. The underlying idea is to copy the source case solution, or to use the rules from the farmer (or a set of farmers) associated to the source case to build the solution. Adaptation knowledge allows to choose the rule to apply among the relevant ones, according to an adaptation context: the user can favorise the rules with conclusion 0 (when the context is not favourable for miscanthus, e.g. because its price is low with regards to traditional crops) or those with conclusion 2 or 1, if the economical context is favourable for miscanthus.

4 SAMM Reasoning Module

In SAMM proptoype, reasoning is based on the two main steps of CBR, retrieval of source cases similar to the target problem, and adaptation of a source case solution to the target problem.

4.1 Retrieving source cases

The retrieval step consists in identifying one or several source cases which can help resolving the target problem. This step is made of three sub-steps: matching of the target problem to source problems, similarity assessment between problems, and selection of a source case. In SAMM prototype, matching the target problem to source cases is a simple vector matching. Similarity assessment between each source problem and the target problem relies on the measures above-defined. Finally various measure combinations can be used for selecting a source case (see Figure 2).

Four retrieval algorithms have been defined. They are specified by the descriptors used for matching problems and the way similarity measures are combined:

– There are two sets of descriptors: the first set, DP, contains the descriptors of the source problem; the second, DI, contains descriptors that index the source case (see section 3.3);
– There are two ways of combining measures: *(i)* the first one first minimizes global distance and then maximizes the number of common descriptors while *(ii)* the second one first maximizes the number of common descriptors and then minimizes the global distance between source cases and target problem.

In the current version of SAMM, the retrieval step selects the source case that minimizes the distance *(i)* or maximizes the number of descriptors *(ii)*. If several cases are returned, then the second measure is used.

In the following, we denote the four retrieval algorithms as RM1 (DP and *(i)*), RM2 (DP and *(ii)*), RM3 (DI and *(i)*), RM4 (DI and *(ii)*).

4.2 Adapting a source case solution

The adaptation step in SAMM is achieved by copying the source solution to the target solution or by transforming the source solution based on the decision rules. Indeed the retrieval step results in a source case which solution (0, 1 or 2) can be directly copied to the the target case. If several source cases are retrieved, the solution is chosen by a majority vote. If none majority can be established a combined solution is proposed, among the following values: 0-1, 1-2, 0-2, 0-1-2.

Various rule sets can be considered for transformation: the rule set of the farmer associated to the retrieved source case, the rule set of a group of farmers (e.g. farmers from the same small agricultural region as the source farmer), the set of all rules. Rules can be applied as soon as each attribute-value (a_i, v_{ir}) in their conditions is similar to one of the target problem descriptors, i.e., for each

```
Matching source case 10385231 and target case 10499801
distance pb between fallows and fallows : 0
distance pb between partially_hydromorphic and drying : 20
distance pb between sharp_shape_AND_convex and sharp_shape : 10
distance pb between foncS_1 and foncS_0 : 20
distance index between fallows and fallows : 0
distance index between sharp_shape_AND_convex and sharp_shape : 10
number of common attributes (DP) : 4
global distance on problem : 12.5
number of common attributes (DI) : 2
global distance on index : 5.0
```

Fig. 2. A example of similarity computing between a source case and a target problem: similarity is computed first on DP then on DI

attribute, the local distance $d_l((a_i, v_{ir}), (a_i, v_{it}))$ is smaller or equal to a given threshold denoted d_{rule}.

When several rules match a target problem it is necessary to select the right rule to be applied. This task is done according to the adaptation context, as said before. Three adaptation algorithms have been therefore implemented. A pessimistic algorithm (ADAPT0) first selects rules with conclusion 0 (see algorithm 1); an optimistic algorithm (ADAPT12) first selects rules with conclusion 2 if it exists rules with conclusion 0, then it selects rules with conclusion 0, and at least rules with conclusion 1 (see algorithm 2); a weighted algorithm (ADAPT3) selects rules with the greatest number of conditions (see algorithm 3). If no rule can be matched with a target problem, then transformation cannot be done and the solution of the retrieved source case is copied into the target solution. If there are several retrieved source cases, the adaptation algorithm is runned for each case and the final result is chosen by a majority vote.

For example, suppose that the retrieved rule set contains the following rules:

R1: $<$ (soil water regime = flood area), $(0) >$
R2: $<$ (field size = middle), $(2) >$
R3: $<$ (drainage = none) $\wedge$ (soil water regime = resurgences and sources), $(0) >$
R4: $<$ (last land-use = fallows) $\wedge$ (soil water regime = wet), $(1) >$

and is used to solve a target problem t with the following descriptors:

    {(size: middle); (drainage: none) ; (soil water regime: wet);
    (excess of water: flood area); (last land-use: fallows)}

The set of matching rules for t, with respect to the distance threshold $d_{rule} = 0$, is {**R1**, **R2**, **R4**}. Algorithm ADAPT0 returns solution 0 (rule **R1** is applied). Algorithm ADAPT12 returns solution 2 (rule **R2**), and the weighted algorithm

---
Algorithm 1: ADAPT0
---
    input : target problem t, set of matching rules S_x with conclusion $x = 0, 1, 2$
    output: problem solution, sol_t
    $sol_t \leftarrow -1$
    if $S_0 \neq \emptyset$ **then**
      | $sol_t \leftarrow 0$
    else if $S_1 \neq \emptyset$ **then**
      | $sol_t \leftarrow 1$
    else if $S_2 \neq \emptyset$ **then**
      | $sol_t \leftarrow 2$
    end

Algorithm 2: ADAPT12
---
    input : target problem c, set of matching rules S_x with conclusion $x = 0, 1, 2$
    output: problem solution, sol_t
    $sol_t \leftarrow -1$
    if $S_0 \neq \emptyset$ **then**
      | **if** $S_2 \neq \emptyset$ **then**
      | | $sol_t \leftarrow 2$
      | **if** $S_2 = \emptyset$ **then**
      | | $sol_t \leftarrow 0$
    else if $S_1 \neq \emptyset$ **then**
      | $sol_t \leftarrow 1$
    else if $S_2 \neq \emptyset$ **then**
      | $sol_t \leftarrow 2$
    end

ADAPT3 selects rule **R4**, with two conditions, and returns solution 1. With the distance threshold $d_{rule} = 5$, the rule **R3** with conclusion 0 can be used, since the distance between values `flood area` and `resurgences and sources` is 5 (attribute `excess of water`). Rule **R3** has two conditions, as rule **R4**; nevertheless, algorithm ADAPT3 first selects rule **R4** which conclusion is 1.

5   Assessing SAMM performance

SAMM prototype has been implemented in java within the Eclipse development environment[3]. Currently, the user can load various rule bases or case bases, and thus can use the model on various territories. The user also can test the algorithms and parameters for the retrieval and adaptation steps, thanks to a configuration panel. He/she can evaluate the proposed solutions *via* the system interface or with output files. Figure 3 represents the various algorithms and bases in SAMM with respect to the CBR general cycle [4].

    Tests have been achieved based on a subset of 72 cases (18 with solution 0 – no miscanthus allocation, 28 with solution 1 – miscanthus allocation, and 26

---
[3] www.eclipse.org

Algorithm 3: ADAPT3
---
    input : target problem c, set of matching rules S_x with conclusion $x = 0, 1, 2$
    output: problem solution, sol_t
    $sol_t \leftarrow -1$, $n_0 = n_1 = n_2 = 0$
    for $x = 0$ to $2 \wedge S_x \neq \emptyset$ **do**
    | $n_x \leftarrow$ average number of conditions for rules in S_x
    end
    if $\exists i, n_i > 0 \wedge n_i > n_j \; \forall j \neq i$ **then**
    | $sol_t \leftarrow i$
    else if $n_2 > 0 \wedge n_2 \geq n_j \forall j \neq 2$ **then**
    | $sol_t \leftarrow 2$
    else if $n_1 = n_0 \neq 0$ **then**
    | $sol_t \leftarrow 1$
    end
---

with solution 2 – miscanthus allocation under conditions). Tests were carried out by extracting from the case base an individual, which is the target problem, while the remaining cases are the source cases. This procedure was repeated for each case in the case base. Various procedures have been tested, based on the combination of various experimental parameters:

– choice of the retrieval algorithm, four modalities, one for each combination of similarity measures, RM1, RM2, RM3, RM4 (see above);
– choice of rule set, in this experiment, we use rules of the farmer associated to the retrieved source case (denoted RSF for *rules of source farmer*)
– choice of the distance threshold between rule conditions and descriptors of the target problem for selecting the rule to be applied ($d_{rule} = 0$ in this experiment);
– choice of the adaptation algorithm, three modalities, one for each algorithm ADAPT0, ADAPT12, ADAPT3 (see above).

Firstly, results obtained at the end of the retrieval step are examined: at this step, the target solution is a copy of the solution (chosen by a majority vote) of the retrieved source cases. The proposed solution is compared to the original solution of the target case. Precision and recall scores are computed for each combination of parameters and each solution (0, 1, or 2). Precision scores are higher than recall ones; this is (partially) due to the fact that some problems got a multi-valued solution when no majority could be etablished (for 1.4 to 9.7% of cases, depending on the algorithm, see Tab. 3). The RM4 algorithm is better both for recall and precision. RM1, RM2, and RM3 obtain similar results, but RM3 has the higher level of cases with a multi-valued solution. RM2 has the lower recall and the lower level of cases with multi-valued solutions. Precision and recall value are globally satisfactory for this first reasoning step, and considering the variety of pairs (attributes,values) in the case descriptions.

Secondly, the effect of adaptation algorithms is assessed based on the parameters described above. Twelve combinations have been defined (4 retrieval
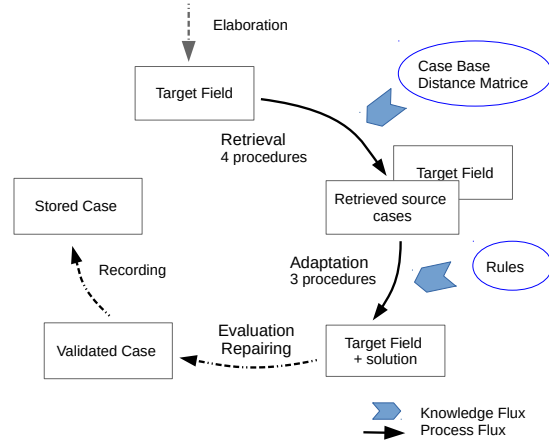
Fig. 3. The CBR cycle in SAMM: dotted lines correspond to currently manual steps

algorithms x 3 adaptation algorihtms x 1 rule set). Two parameters are measured, the adaptation level, i.e. the number of target cases with an adapted solution, and the distribution of solution values (0, 1, 2 or multi-valued) in adapted cases.

Adaptation levels vary from 40 to 50% using the rules of the source farmer (RSF). The best level (50%) is given by algorithm RM3, which first optimises global distance $d_g(s,c)$ on set DI (index descriptors), and then the number of common attributes $n_{DI}(s,c)$. This is to be linked to the fact that case index parameters have been defined based on farmer explanations about farm fields; rule conditions have also been defined on the same basis. On the contrary, RM2 has the lowest adapatation level (and also the lowest recall and precision levels): it first optimises $n_{DP}(s,c)$, i.e. the number of common attributes in DP (problem descriptors).

	RM1	RM2	RM3	RM4
recall	55.65	52.48	54.32	56.72
precision	59.70	59.11	60.37	70.39
% multi-valued solution	4.17	1.39	9.72	5.56

Table 3. Results for the four retrieval algorithms

Table 4 shows the distributions of computed solutions for the 12 algorithm-parameter combinations. Results show that pessimistic algorithm ADAPT0 allows to favour solution 0 especially for RM1 and RM3 (the level of solution 0 is 55-56%). Optimistic algorithm ADAPT12 favours solution 2 for RM1, RM2,

and RM3 (the level of solution 2 is between 47 and 52%; it is lower for RM4). Weighted algorithm ADAPT3, which also favours solution 2, gives results similar to these of ADAPT12, but the level of solution 1 can be higher (from 19 to 26% for RM4). Solution 1 is seldom chosen for all algorithm combinations (between 8 and 26%): this is partially due to the algorithm structures (especially ADAPT0 and ADAPT12) but also to the low number of rules with conclusion 1 in the rule base (see Section 3.2). Besides, multi-valued solutions are not many: this shows that even when several source cases are retrieved, and then several rules can be used, the adaptation algorithms allow to compute a dominant solution.

	ADAPT0	ADAPT12	ADAPT 3	ADAPT0	ADAPT12	ADAPT 3
	RM1			RM2		
solution 0	0.55	0.39	0.39	0.41	0.34	0.31
solution 1	0.09	0.09	0.09	0.14	0.14	0.17
solution 2	0.36	0.52	0.52	0.41	0.48	0.48
multi. sol.	0	0	0	0.03	0.03	0.03
	RM3			RM4		
solution 0	0.56	0.42	0.42	0.41	0.42	0.35
solution 1	0.08	0.08	0.08	0.19	0.19	0.26
solution 2	0.33	0.47	0.47	0.34	0.35	0.35
multi. sol.	0.03	0.03	0.03	0.06	0.03	0.03

Table 4. Adaptation: distribution (%) of computed solutions by adaptation algorithms, for each retrieval algorithm, using rules of the source farmer

These results highlight that various algorithm/parameter combinations can be used, according to the user objectives: if he or she wants to forecast the allocation of miscanthus in a negative context, he or she will use the RM1-ADAPT0-RSF or RM3-ADAPT0-RSF combinations; for a positive context (e.g. good economic conditions), he or she will use the ADAPT12-RSF combination with either RM1, or RM3. Finally, to be close to current farmer choices, it will be better to use the RM4-ADAPT3-RSF combination.

This experiment also shows the central role of the user, who has to choose algorithms and parameters, and to examine results step by step: retrieved cases, available rules, proposed solutions. Rules in particuler can be analysed to highlight the field characteristics that are important wrt the farmer decision.

6   Conclusions and perspectives

We have presented a CBR system, SAMM, which aim is to forecast the allocation in french farms of a new biomass energy perenial crop, the miscanthus. The system includes a case and a rule base, that have been modeled on the basis of farmer interviews. Similarity measures have been defined and various algorithms for the retrieval and adaptation steps have been implemented. The

system has been tested on a subset of 72 cases, each case representing a farm field and its allocation potential for miscanthus. Results have shown the various behaviours of the system, and thus have highlighted the predominant role of the user, who chooses how to combine algorithms and parameters according to his/her objectives.

According to agronomist point of vue, SAMM modeling and implementation has allowed two advances: *(i)* the formalisation of farmer decision rules, based on complex farm surveys; *(ii)* the contruction of various scenarios for forecasting miscanthus allocation. However, there is a difficulty: since cases are described with few attributes, case source retrieval relies on a weak matching (target and source problems have only 2 or 3 common descriptors). This underlines the importance of the rules used in the adaptation step, which should be carefully chosen. Furthermore, farmer reasoning includes various spatial scales (the farm fields, the field clusters, the farm). Description of rules and cases could be completed to include these various information scales and thus obtain more reliable results. Besides, the problem elaboration step should be implemented, based on existing work [18], to make the prototype usable on a larger territory.

Rules can also be used to explain the behaviour of the system and the proposed solution of a target problem. Actually, our work deals with explanation based reasoning. Explanation is a most interesting notion for exploring decisional knowledge, especially when the context is evolving, as it is when a new crop is introduced in farms, with new characteristics and specific issues. Furthermore, the complexity of farmer reasoning processes pleads for a CBR system with capabilities for interacting with the user rather than only capabilities for forecasting. The current prototype is useful to build new loops of survey, modelling and implementation steps, which will lead to further systems with actual capabilities in interacting with the user, and in forecasting or decision-making.

To help the user, the system could also include text elements from the farmer interviews, or drawings they made of their fields, that would complete the case descriptions and rules. Such improvements are proposed in knowledge approaches dealing with knowledge provenance [20]. The underlying idea is both to help the user and to make the system maintenance easier. Finally, an interesting prospect for our work is to directly associate resolution and training on the case base, allowing the concerned stakeholders to appropriate the modeling process, and to share knowledge between them.

References

1. Hellmann, F., Verburg, P.: Spatially explicit modelling of biofuels crops in europe. Biomass & Bioenergy **35** (2008) 2411–2424
2. Martin, L., Wohlfahrt, J., Le Ber, F., Benoît, M.: Perennial biomass crop cultivation and its territorial patterns A case-study of miscanthus in Côte-d'Or (Burgundy, France). Espace Geographique **2**(41) (2012) 133–147

3. Rounsevell, M.D., Arneth, A.: Representing human behaviour and decisional processes in land system models as an integral component of the earth system. Global Environmental Change **21**(3) (2011) 840 – 843 Symposium on Social Theory and the Environment in the New World (dis)Order.

4. Aamodt, A., Plaza, E.: Case-Based reasoning: Foundational Issues, Methodological Variations, and System Approachs. AICOM **7**(1) (March 1994) 39–59

5. Leake, D.B., ed.: Case-Based Reasoning. Experiences, Lessons, & Future Directions. AAAI Press / The MIT Press (1996)

6. Riesbeck, C.K., Schank, R.C.: Inside Case-Based Reasoning. Lawrence Erlbaum Associates, Hillsdale, New Jersey (1989)

7. Brüninghaus, S., Ashley, K.D.: The Role of Information Extraction for Textual CBR. In Aha, D.W., Watson, I., eds.: Case-Based Reasoning Research and Development. Volume LNCS 2080. Springer (2001) 74–89

8. Adeyanju, I., Wiratunga, N., Lothian, R., Sripada, S., Lamontagne, L.: Case Retrieval Reuse Net (CR2N): An Architecture for Reuse of Textual Solutions. In McGinty, L., Wilson, D.C., eds.: Case-Based Reasoning Research and Development. Volume LNCS 5650. Springer (2009) 14–28

9. Le Ber, F., Napoli, A., Metzger, J.L., Lardon, S.: Modeling and comparing farm maps using graphs and case-based reasoning. Journal of Universal Computer Science **9**(9) (2003) 1073–1095

10. Benoît, M., Rizzo, D., Marraccini, E., Moonen, A.C., Galli, M., Lardon, S., Rapey, H., Thenail, C., Bonari, E.: Landscape agronomy: a new field for addressing agricultural landscape dynamics. Landscape Ecology **27**(10) (2012)

11. Du, Y., Liang, F., Sun, Y.: Integrating spatial relations into case-based reasoning to solve geographic problems. Knowledge-Based Systems **33** (2012) 111–123

12. Holt, A., Benwell, G.: Case-Based Reasoning and Spatial Analysis. Journal of the Urban and Regional Information Systems Association **8** (1996) 27–36

13. Branting, L.K., Hastings, J., Lockwood, J.: CARMA: A Case-Based Range Management Advisor. In: Proceedings of The Thirteenth Innovative Applications of Artificial Intelligence Conference (IAAI-2001), Seattle, Washington. (2001)

14. Girard, N., Bellon, S., Hubert, B., Lardon, S., Moulin, C.H., Osty, P.L.: Categorising combination of farmers' land use practices: an approach based on examples of sheep farms in the south of france. Agronomie **21** (2001) 435–459

15. Bosch, O.J.H., Gibson, R.S., Kellner, K., Allen, W.J.: Using case-based reasoning methodology to maximise the use of knowledge to solve specific rangeland problems. Journal of Arid Environments **35** (1997) 549–557

16. Schank, R.C., Kass, A., Riesbeck, C.K., eds.: Inside Case-Based Explanation. LEA, Hillsdale, New Jersey (1994)

17. Leake, D., McSherry, D.: Introduction to the Special Issue on Explanation in Case-Based Reasoning. Artificial Intelligence Review **24**(2) (2005) 103–108

18. Rizzo, D., Martin, L., Wohlfahrt, J.: Miscanthus spatial location as seen by farmers: A machine learning approach to model real criteria. Biomass and Bioenergy **66** (2014) 348–363

19. Carbonell, J.G.: Learning by analogy: Formulating and generalizing plans from past experience. In Michalski, R.S., Carbonell, J.G., Mitchell, T.M., eds.: Machine Learning. Volume 1 of Symbolic Computation. Springer (1983) 137–161

20. da Silva, P., McGuinnes, D., McCool, R.: Knowledge provenance infrastructure. IEEE Data Eng. Bull. **25**(2) (2003) 179–227