Modeling farmers' choice of miscanthus allocation in farmland : a case-based reasoning model

Laura Martin, Florence Le Ber, Julie Wohlfahrt, Géraldine Bocquého, Marc Benoît

International Environmental Modelling and Software Society (iEMSs)
2012 International Congress on Environmental Modelling and Software
Managing Resources of a Limited Planet, Sixth Biennial Meeting, Leipzig, Germany
R. Seppelt, A.A. Voinov, S. Lange, D. Bankamp (Eds.)
http://www.iemss.org/society/index.php/iemss-2012-proceedings

Modeling farmers' choice of miscanthus allocation in farmland: a case-based reasoning model

Laura Martin[a], Florence Le Ber[b], Julie Wohlfahrt[c], Géraldine Bocquého[d], Marc Benoît[e]

[a.] laura.martin@mirecourt.inra.fr
[b.] florence.leber@engees.unistra.fr
[c.] julie.wohlfahrt@mirecourt.inra.fr
[d.] gbocqueho@grignon.inra.fr
[e.] marc.benoit@mirecourt.inra.fr

Abstract: The spatial extension of perennial biomass crop, like miscanthus, seems to be unavoidable to face the decrease of fossil fuel. However, the risk of a food / non food competition due to land use change has to be anticipated. Several models of biomass crops allocation have been already performed. Most of these models simulate large-scale allocation processes, taking into account numerous biophysical variables but only few true-to-life human variables. In this paper, we present a modeling framework of miscanthus allocation in farmland. We use a case based reasoning model in order to compute both biophysical and human variables. An *Ad hoc* similarity measures framework and the comparison of two modelling techniques are presented. First results of one application based on a french case study are discussed. They show the necessity to take into account stakeholders' knowledge of miscanthus allocation process in the modelling.

***Keywords**:* artificial intelligence; decision-making support; miscanthus; modeling; land use

1 INTRODUCTION

To face the decrease of fossil energy supplies, new renewable energy resources like perennial biomass crops are of a great interest (R.E.D., 2009). Their spatial extension and allocation seem then unavoidable, like anticipating global issues as food / non-food competition (Karp, Richter, 2011). Several land-use change models deal with biomass crops allocation (Hellmann, Verburg, 2008; Lovett *et al*., 2009). Most of these models simulate large-scale allocation processes, taking into account numerous biophysical variables but only few true-to-life human variables. Thus, our aim is to model farmers' allocation choice regarding miscanthus, as a complex agricultural management system, coupling social, technical or environmental variables to assess biomass spatial distribution.
As, coupling human and biophysical variables in a modeling framework raises knowledge acquisition and knowledge integration methodological questions, we propose to model biomass crop allocation relying on the case-based reasoning model (Riesbeck, Schank, 1989; Aamodt, Plaza, 1994). The choice of this model is explained and tested in a case (Burgundy biomass cooperative). This work is part of the FUTUROL project which deals with industrial process of ligno-cellulosic biomass resources.
This article presents successively the case based reasoning method, the first application to miscanthus allocation modeling, and focus the results on two main scientific questions: (i) how to retrieve a similar case, (ii) how to reuse retrieve

case's solution to predict miscanthus allocation? We close this paper through a short conclusion on the model status in human decision making.

2 MATERIAL AND METHOD

2.1 Case-based reasoning theory and assets

Case-based reasoning (CBR) is a problem solving paradigm based on analogy reasoning. It belongs to Artificial Intelligence sciences. CBR consists in solving new problem by using the solution of similar old problem already solved (Riesbeck, Schank, 1989). For instance in land design, a new problem can be the prospective (potential?) miscanthus allocation into one farmland of a small region, and an old problem is so the miscanthus allocation observed in similar farmlands.
A *Case* corresponds to a problem-solving episode represented by the pair *Problem-Solution* and by all the information related to the path dependency between the *Problem* (a farmland) and its *Solution* (the miscanthus allocation into the farmland).
CBR process consists in solving a *Target problem* - a new problem - by using a *Case Base* (solved cases) according to the following four stages (cf. figure 1): 1. retrieve the most similar case - named a *Source case* - to the *Target problem* by similarity measures between problems, 2. reuse the *Solution* of the *Source case* by inference processes and adaptation knowledge, 3. revise the *Target solution* (the inferred *Solution*) if necessary and 4. retain the *Target case* and its problem-solving episode as a new *Case* into the *Case Base* (Aamodt, Plaza, 1994, Watson, Marir 1994). In CBR, the two major steps are stage 1 and stage 2. For instance in stage 2, if the miscanthus is allocated on maize plots in *Source case*, CBR can either use the *Solution source* to allocate miscanthus on maize plots, or adapt the *Solution source* to the *Target case* constraints, like flood risk area which one prevent the allocation of miscanthus on maize plots located in such area for harvesting reasons (Valmi-Dufour et al., 2012).

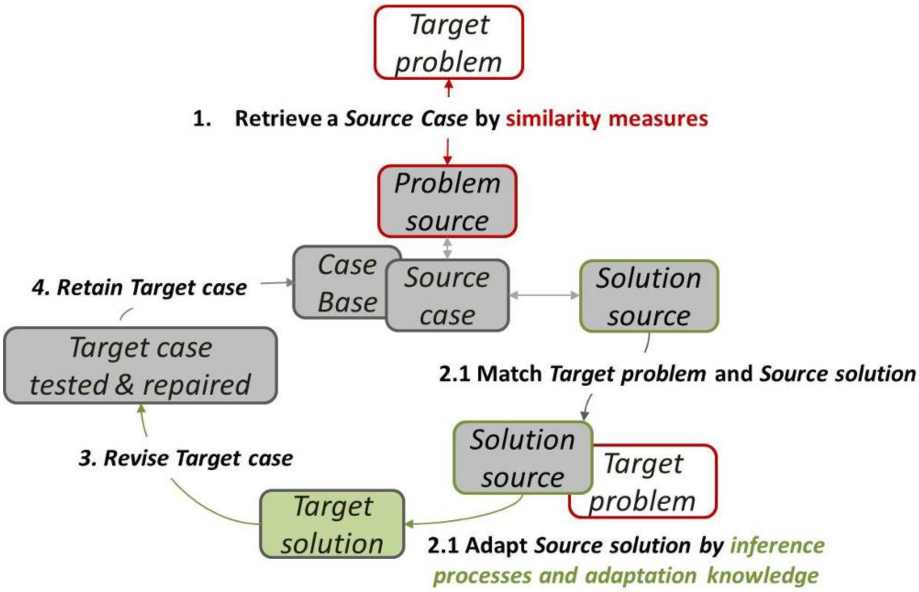


Figure 1: the CBR process (adapted from Aamodt, Plaza, 1994)

The major asset of CBR is to be able to model complex mechanisms like environmental ones (Mota *et al*, 2008) without the necessity to fully understand driving mechanisms (Yunyan *et al*., 2010). Indeed, the analogy reasoning is able to solve problems by (with?) few data. This asset is perfectly adapted to perennial biomass allocation issues which ones are too recent to be fully understood, or represented by statistics and by stakeholders' decision rules and interactions. Therefore CBR model can be an alternative to agent-based models as its

reasoning is more global and less distributed: CBR is not based on agent reasoning but on *Cases*, i.e. on the transposition of current allocation practices but not on simulated interactions and decision rules obtained by simulations with stakeholders (Matthews *et al*., 2007). The ability of CBR to take into account different types of knowledge (Leake, 1996) gives also the opportunity to integrate heterogeneous data like biophysical and human variables. Mainly used in industrial and medicinal problematic, case-based reasoning is also used in Land Use Change Science, even if there are still few results published (Yunyan *et al*., 2010).
Thus, we consider the use of CBR to model the allocation of perennial biomass crop as an innovative approach to integrate complex local stakeholders' knowledge. The following sections present one specific application case: the CBR miscanthus model.

2.2 Case-based reasoning to model miscanthus allocation

2.2.1 Study area

The CBR Miscanthus model (CBRMM) is based on a case study located in Burgundy (Côte d'Or), a region area situated in the east of France, where substantial process of miscanthus implantation is currently observed (cf. figure 2). As a matter of fact, in this area, european subsidies are given to farmers to support miscanthus.

This case study includes several research teams from the FUTUROL project, gathered together to mutualize research progresses and results.
Therefore, our application is based on pooled data coming from two INRA (French National Institute for Agricultural Research) research teams: Public Economy and SAD-ASTER. Both research teams carried out respectively 111 individual farm surveys and 10 comprehensive interviews of farmers, in 2010 and 2011 (cf. table 1).

Table 1: samples of survey and comprehensive interviews

	Number of farmers by main activity			
Survey kinds	Cereal grower	Cattle breeder	Other	**Total of farmers**
Individual farm's surveys	85	22	4	**111**
Individual comprehensive interviews	6	4	0	**10**

Figure 2: localisation of the study area in France

Whereas survey data are mainly used to fill the values of the *Case description* (cf. table 2), comprehensive interviews are used to select the attributes of the *Case description* and more globally, are used to build domain knowledge and adaptation rules for retrieve and adaptation stages (cf. figure 1).
Indeed, a comprehensive interview differs from a survey because it includes no leading questions (Kaufmann, 1996). It is adapted to catch all factors influencing farmers' choice from diverse kinds (social, technical) and from diverse degree of complexity (mono-factors and multi-factors). The interview is recorded, fully transcribed and analyzed, enabling to catch decision rules (cf. table 2, table 7) and driving factors explaining both the miscanthus adoption and its allocation into farmland - like the distance to the farm-stead, farmer's perception of biophysical and spatial farmland features, the cropping plan etc. (Martin *et al*., 2012).

2.2.2 Case description

A Case is represented by objects which are described by a set of "attribute-value" (Bergmann *et al*., 1998). Some objects belong to the *Problem* part of the *Case* and others to the *Solution*.
In the CBRMM, *Problem* corresponds to the driving factors expressed by farmers for which the modeler has selected correspondent attributes that best describe them. Exactly, *Problem* corresponds to a farmer and farm features coupling both biophysical and human attributes. It is composed of four attributes groups linked to socio-technical processes - farmer's attributes, cropping plan - and to more biophysical processes - farm biophysical and spatial farmland features (cf. table 2).
Solution corresponds to miscanthus allocation practices and miscanthus plot features (cf. table 6, part 3.2.2).

Table 2: Humans *(in red) and biophysical attributes (in orange) of the Problem*

Farmer's attributes
main activity and land tenure system of plots
farmer's allocation rules
farmer's perceptions of biophysical and spatial farmland features
farmer's perceptions of miscanthus

Data source : Comprehensive interviews **Number of cases : 10/111**

Cropping plan	**Farm biophysical features**
usable agricultural area (ha) - UAA	textural soil classification
arable land area (ha & % of UAA)	area (ha) without/with slope (from 5 to 10%)
land under permanent grass area (ha & %)	**Spatial farmland features**
set-aside area (ha & %)	number of plots and area (ha) located at different distances to the farm-stead
permanent crops (ha & %) - e.g. vineyard	number of plots from different size
perennial crops area (ha & %) - e.g. miscanthus	number of plots and area (ha) located near forests, rivers and houses

Data source : Surveys & geographically referenced data / Number of cases: 111

One of the key problems of CBR frameworks is finding similar cases in the *Case Base*. The choice of similarity measure is important for the success of the adaptation process. Even if several similarity measures are commonly used in CBR like nearest neighbor and ExpertClerk Median algorithm, Lucene retrieval method in jCOLIBRI (a CBR open source framework, R. Garcia, 2008), the difficulty is the selection of the attributes (and weight) to compare. As they need to be well adapted to the problem-solving issue, similarity measures must be adapted to each CBR application and cannot be completely generic or imported from other CBR frameworks.
For the CBRMM, an *Ad hoc* similarity measures and adaptation framework have been chosen.

3 RESULTS

3.1 *Ad hoc* similarity framework

To define the *Ad hoc* similarity measures framework, we assume that similar farm management and biophysical constraints of farmland enable analogue farmers' choices regarding crops allocation. The comparison of the *Target* case (appointed *tgCase*) and cases of the *Case Base* (*bsCases*) is based on a combination of three of the four components of the *Problem part*: cropping plan, farm biophysical features and spatial farmland features.
We first detail the similarity measure of cropping plans, then the similarity measure of soils.

3.1.1 Similarity measure of cropping plan

Retrieve similar cropping plan is a major step, considering it drives farmers' choices about crop dynamics and miscanthus allocation. To compare cropping plans, we compare the crops proportions in each farm. We assume that a similar cropping plan indicates a similar crop production activity of the farm, similar cropping schedule and work calendar, close crop rotations and similar crop requirements (e.g. water and soil).
To compare cropping plans, we use two indexes. The first one compares the proportions of common crops between *tgCase* and the *bsCases* plans. The second one compares the proportions of non-common crops.
As our goal is to retrieve not only a similar cropping plan but a similar crop allocation management, we use weighted coefficient for computing the two indexes, as fallow for non-common crops (second index):

$$nonCommonCropsIndexe = \sum_{i=1,\, j=1}^{n,m} \left(\left| nCrop\_T\arg et(i) + nCrop\_CB(j) \right| \times wc(i,\, j) \right) \qquad (1)$$

Where:
n = number of crops (i) only produces by *bsCase* and not by *tgCase*
m = number of crops (j) only produce by *tgCase* and not by *bsCase*
nCrop_tgCase(i) = proportion of crops (i) in the *tgCase* cropping plan
nCrop_*bsCase* (j) = proportion of crops (j) in the *bsCase* cropping plan
wc (i,j)= weighted coefficient

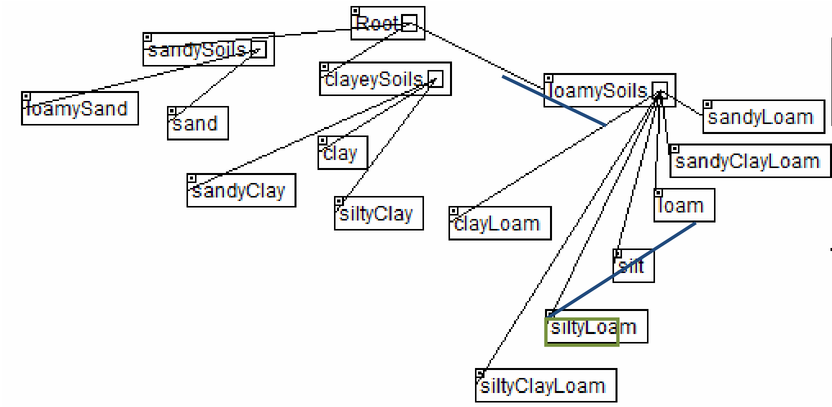
The aim is to strengthen the retrieve process to similar crops requirement management by considering "more similar" two crops having close agronomical and/or technical requirements. For instance, if the *tgCase* produces maize, thanks to weighted similarity measures we retrieve both *bsCases* which produce maize and *bsCases* which produce similar crops regarding moisture content requirements (e.g. soya, miscanthus) (cf. table 3).

Table 3: weighted coefficients values

wc Values	Level of similarity / dissimilarity with crops of *Target problem*
0 < wc < 1	similar allocation requirements with close cropping systems
wc = 1	dissimilarity of allocation requirements
wc > 1	dissimilarity of allocation requirements with different farming management (e.g. cropping system, farm activity)

3.1.1 Similarity measure of soil types

For our application, we would only use the information about soil texture to account for biophysical farm features attributes (cf. table 2). First we compare the proportion of soil texture between *tgCase* and *bsCases*. Then, according to the procedure done for cropping plan similarity measures, we established weight to account for the proximity between different kinds of soils. To compare soil textures we built a taxonomy (a hierarchical set of concepts) based on the FAO soil textural classes, where the final sheets correspond to textural classes and where upper nodes correspond to more general textures (cf. figure 3). The similarity between soil types is expressed by a path length to a common parent. The distance is calculated by the number of nodes between two soil textures according to figure 3. The example of one path length from the *Target case* (in green) to the *Case Base* (in red), corresponding to "one different parent", is represented in figure 3.



Hierarchy positions of *Target* and *Case Base*	wc Value
common parent	1,5
one different parent	2
n different parents	n+1

Figure 3: textural soil hierarchy and weighted coefficient values.

The similarity measure of soil types can be calculated as fallow:

$$soilSimilarity = \sum_{i=1,j=1}^{n,m} \left(\left| soil\_T\arg et(i) + soil\_CB(j) \right| \times wc(i,j) \right) \qquad (2)$$

Where:
N = number of soil textural classes of *tgCase* farmland
m = number of soil textural classes of *bsCase* farmland
soil_*tgCase*(i) = proportion of soil textural classe (i) in *tgCase* farmland
soil_*bsCase*(j) = proportion of soil textural classe (j) in the *bsCases* farmland
wc (i,j)= weighted coefficient

For spatial features, similarity measures have been computed by comparing the proportion of each spatial feature (cf. table 5) between *tgCase* and *bsCases.*

3.2 Application first results

3.2.1 Retrieve process results

The aim of our application is to retrieve the similar *Case* to the *Target case E4*, for which one, *Problem* and *Solution* have been caught by survey and comprehensive interview (cf. table 5, table 6, table 7).
The final score measuring the global similarity between *tgCaseE4* and *bsCases* is the sum of the different measures for each *Case*. The most similar *Case* regarding *tgCaseE4* is the *Case13* described as fallow.

Table 5: *Target problem* and *Source Case* descriptions, and dissimilarity results

Cropping plan	***tgCaseE4***	***Case13***	***Dissimilarity***
proportion of set-aside over UAA (%)	*5,6*	2	3,6
maize proportion of arable land (%)	*10,9*	14,5	-3,6
Total (%)	*5,6*	2	3,6
Spatial farmland features	***tgCaseE4***	***Case13***	***Dissimilarity***
number of field blocks	25	45	-20,0
total area (ha) of distance plots to the farm-stead ≤ 1 km	95	0	95,0
tot.area of distance plots to farm-stead: >1 km,10 km<	60	211	-151,0
total area (ha) of distance plots to the farm-stead ≥ 10 km	25	0	25,0
number of plots located near woodland	2	8	-6
number of plots located near rivers	12	5	7

3.2.2 Adaptation results and validation

In order to adapt the *Solution source* to *tgCaseE4*, several more or less complex techniques can be used (Watson, Marir, 1994). As the *Solution* of *tgCaseE4* is known, we compared 2 adaptation techniques by the validation of adaption results.
We first tested a "null adaptation" technique. Crucial differences between the inferred and real *Solutions* were pointed out as the surface of miscanthus plots (cf. table 6). This situation reveals differences of practices between both farmers and shows us the necessity to take into account dissimilarity between *Cases* to infer *Solution*. Another "simple" technique could be to adjust attribute-value pairs according to the dissimilarity (cf. table 6) as fallowing adaptation rules:
r1. if "*tgCaseE4* number of plots < *Case13* number of plots" then "*tgCaseE4* number of miscanthus plots < *Case13* number of miscanthus plots"
r2. if "*tgCaseE4* distance of most far-off plots > *Case13* distance of most far-off plots" then "*tgCaseE4 distance of* miscanthus > *Case13 distance of* miscanthus"
r3. if "*tgCaseE4* number of plots located near river > *Case 13* ones" then *tgCaseE4* miscanthus will be allocated near a river.

Table 6: adaptation results and validation in red incorrect solution

	real *Solution* of the *Target* case E4		***Solution* from null adaptation (⇔case 13 sol)**			***Solution* from simple adaptation**	
numb. of miscanthus plots	2		3			2	
area of plots (ha)	15	5	**3,2**	**1,21**	**1,81**	**2,63**	**1,24**
soil type	Clay		Clay loam (CL)	*CL*	CL	CL	CL
land tenure system	owner occupancy		owner occupancy			owner occupancy	
past 3 years covers	maize	set-aside	set-aside	maize	maize	set-aside	maize
dist. to farm-stead	20 km		**7 km**			**> 10 km**	
slope pourcentage	0		0			0	
flood-risk of plots	yes		no			no	
neighborhood feature	woodland, river		river	woodland	river	river	

Results show that the simple adaptation techniques are not sufficient to adapt correctly the *Solution source*. More elaborate methods as integrating farmers' decisions rules (cf. table 7) should be applied, as they explain allocation practices of *tgCase*. A feedback can also be necessary to change similarity measures.

Table 7: *Target problem* description

Farmer's attributes	
miscanthus allocation decision rule 1	allocation in nitrate-vulnerable zone
miscanthus allocation decision rule 2	allocation in far-off plots
miscanthus allocation decision rule 3	flood risk of plot
miscanthus allocation decision rule 4	good agronomical value of the plot
management decision rule	compensate sugar beet production stopping
perception of miscanthus	crop friendly environmental
perception of farmland textural soils	good agronomical value of textural soils
perception of spatial farmland feature	transport costs constraint of far-off plots

4 DISCUSSION AND CONCLUSION

In this paper, we described a preliminary CBR application to predict allocation dynamics of miscanthus in farmland. An *Ad hoc* similarity measures framework has been built to retain the most similar *Cases* regarding land use change management of farmer and its ability to allocate miscanthus in farmland.
At the present time, similarity measures and adaptation process are based on three attributes groups: copping plan, biophysical farmland features and spatial

farmland features. But as we saw in part 3.2.2, it is necessary to take into account farmers' rules for the adaptation process and for similarity measures. Thus, our future work will consist to enrich our model by farmers' decision rules and perceptions, in an iterative way. Even if we do not build the CBRMM in a participative way, we are going to use farmer's choices to calibrate adaptation rules and to validate them by feedbacks. A second period of interviews with farmers is planned to catch their adaptation practices, according to different scenarios (built by the searcher beforehand). On the other hand, similarity level between crops has been defined according to crops requirements and major features that broadly influence farmer allocation rules. To increase the validity of the model to local application, it could be interesting to use the observation of local cropping system for several years. The use of Terruti data (Mari, Le Ber, 2006) can be a work perspective.

To conclude, case based reasoning provides an interesting opportunity to integrate various data, like survey data but also like stakeholders' rules and choices. More than being an alternative to model land use change, we hope that the use of CBR could also be an efficient way to fully understand current and future practices of biomass allocation, in order to anticipate the food/ non food competition risk.

ACKNOWLEDGMENTS

This work was funded by FUTUROL project and OSEO. Thank you to them for this support. Thank you also to all farmers and other local stakeholders interviewed.

REFERENCES

Aamodt, A., Plaza, E. Case-based reasoning - foundational issues, methodological variations, and system approaches. *Ai Communications*, Vol. 7, p39-59, 1994.

Dufour-Lussier, V., Le ber, F., Lieber, J., Martin, L. Adapting Spatial and Temporal Cases. ICCBR 2012

European Commission. Renewable Energy Directive (Directive 2009/28/EC) of the European Parliament and of the Council, dated of the 23 April 2009.

Gaucherel, C., Griffon, S., Houet, T., Misson, L. Combining process-based models for future biomass assessment at landscape scale. *Landscape Ecology*, Vol. 25, p. 201-215, 2009.

Hellmann, F., Verburg, P. Spatially Explicit Modelling of Biofuels Crops in Europe. *Biomass & Bioenergy,* Vol. 35, p. 2411-2424, 2008.

Karp, A., Richter, G. M. Meeting the challenge of food and energy security. *Journal of Experimental Botany*, Vol. 62, 2011.

Kaufmann, J.-C. L'entretien compréhensif. Paris, Nathan, 1996.

Leake, D. B. CBR in context: The present and future. *Case-based reasoning: Experiences, lessons, and future directions*, p 3–30, 1996.

Lovett, A.A., Sünnenberg, G.M., Richter, G.M., Dailey, A.G., Riche, A.B., Karp, A. Land use implications of increased biomass production identified by GIS-based suitability and yield mapping for Miscanthus in England. *BioEnergy Research*, Vol.2, p. 17-28, 2009.

Mari, J-F., Le Ber, F. Temporal and spatial data mining with second-order hidden markov models. *Soft Comput*, Vol 10, p 406–414, 2006.

Martin, L., Wohlfahrt, J., Le Ber, F., Benoît, M. Perennial biomass crops allocation: a French case study regarding miscanthus. *L'Espace Géographique*, 2012 (submitted).

Mota, J. S., Câmara, G., Escada, M. I. S., Bittencourt, O., Fonseca, L. M. G., Vinas, L. Case-based reasoning for eliciting the evolution of geospatial objects, *Springer-Verlag*, p 405-420, 2009.

Riesbeck C. K., Schank R. C. Inside Case-Based reasoning. Lawrence Erlbaum Associates Publishers, Hillsdale, New Jersey, 1989.

Watson, I., Marir, F. Case-based reasoning: A review. *The Knowledge Engineering Review*, Vol 9 , pp 327-354, 1994.

Yunyan, D., Wei, W., Feng, C., Min, J. A case-based reasoning approach for land use change prediction. *Expert Systems with Applications*, Vol 37(8), p 5745-5750, 2010.