Miscanthus spatial location as seen by farmers: a machine learning approach to model real criteria

RIZZO Davide [a,b] , MARTIN Laura [c] , WOHLFAHRT Julie [d]

[a] INRA SAD-ASTER, 662 Avenue Louis Buffet F-88500 Mirecourt, France
E-mail: ridavide@gmail.com
[b] corresponding author: Telephone. +33 (0)329385507, fax +33 (0)329385519

[c] INRA SAD-ASTER, 662 Avenue Louis Buffet F-88500 Mirecourt, France
E-mail: laura.martin@mirecourt.inra.fr

[d] INRA SAD-ASTER, 662 Avenue Louis Buffet F-88500 Mirecourt, France
E-mail: julie.wohlfahrt@mirecourt.inra.fr

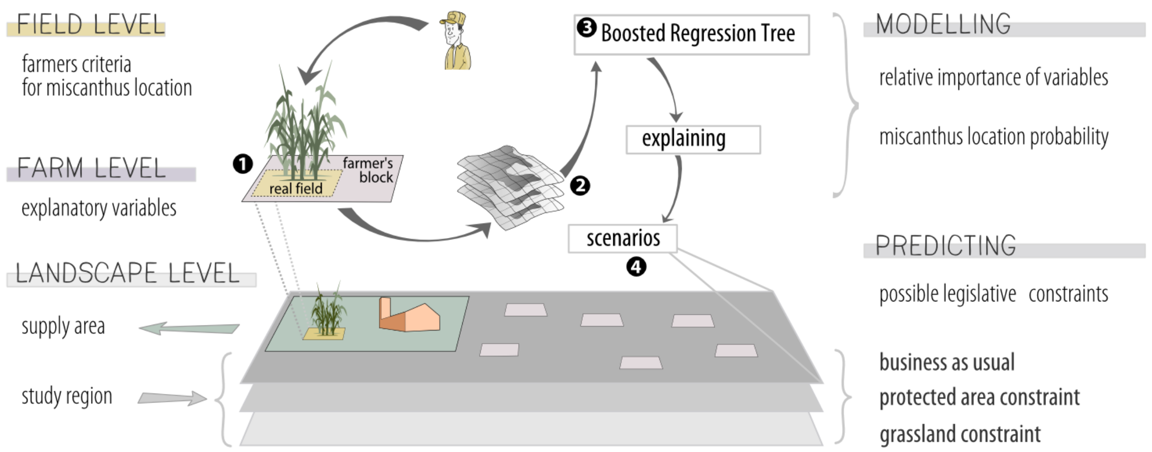
Type of paper: research paper

Date of the manuscript: February 22nd 2014

Word manuscript count: 6296

Please cite this article in press as: **Rizzo D, et al.**, Miscanthus spatial location as seen by farmers: A machine learning approach to model real criteria, *Biomass and Bioenergy* (2014), http://dx.doi.org/10.1016/j.biombioe.2014.02.035

Graphical Abstract



ABSTRACT

Miscanthus is an emerging crop with high potential for bioenergy production. Its effective sustainability depends greatly on the spatial location of this crop, although few modelling approaches have been based on real maps. To fill this gap, we propose a spatially explicit method based on real location data. We mapped all of the miscanthus fields in the supply area of a transformation plant located in east-central France. Then, we used a boosted regression tree, machine learning method, to model miscanthus presence/absence at the level of the farmer's block as mapped in the French land parcel identification system. Each of these modelling spatial units was characterised on agronomical, morphological and contextual variables selected from in-depth spatially explicit farm surveys. The model fostered a two-fold aim: to assess the farmers' decision criteria and predict miscanthus location probability. In addition, we evaluated the consequence of possible legislative constraints, which could prevent the miscanthus to be planted in protected areas or in place of grasslands. The small and complex-shaped farmer' blocks that were predicted by our model to be planted with miscanthus were also characterised by their great distance from the farm and the roads. This kind of result could provide a different perspective on the definition of "marginal land" by integrating also the farm management criteria. In conclusion, our approach elicited real farmers' criteria regarding miscanthus location to capture local specificities and explore different miscanthus location probabilities at the farm and landscape levels.

KEYWORDS

bioenergy crop, land parcel identification system (LPIS), landscape agronomy, Boosted Regression Trees (BRT), France, marginal land

ABBREVIATIONS

AUC-ROC, area under the receiver operating characteristic; BRT, boosted regression tree; EJ, exajoules; IACS, integrated administration and control system; IGN, Institut national de l'information géographique et forestière; LPIS, Land parcel identification system; NUTS, Nomenclature of Territorial Units for Statistics

1. INTRODUCTION

Biomass feedstock is the first source of renewable energy worldwide and its availability for bioenergy production will be a major issue for the future decades. The bioenergy contribution to the primary energy supply in 2008 – valued at 492 exajoules (EJ) – has been estimated at 10.2% compared to 12.9% of the total renewable energy contribution [1]. The use of biomass for bioenergy is expected to increase further as we will face the energetic transition that fosters replacing fossil fuels by renewable resources [2,3]. The technical potential of biomass energy crops for 2050 is estimated in approximately 96 EJ/yr [4] with expert-based potential deployment levels being assessed in the range of 100 to 300 EJ/yr [5]. Bioenergy can be produced from a variety of biomass residues, short-rotation forest plantations, energy crops and organic wastes. Although agricultural and forestry co-products can provide the major share of the biomass feedstock supply [6], a substantial portion of the demand is expected to be met by cultivating dedicated energy crops [7]. In particular, perennial energy crops have been shown to be good candidates for bioenergy production [8–10] and to have a relatively low environmental pressure compared to annual crops [11,12]. These crops could contribute to the sustainable intensification of farming systems and landscape structure that can provide multiple ecosystem services [5,13–17]. Moreover, perennial crops can reduce cultivation costs because they have no need for annual planting and have reduced tillage requirements [18]. Additionally, these contributions are also key advantages to meet the sustainability requirements defined by the European Union Renewable Energy Directive (2009/28/EC).

Cultivating dedicated energy crops raises, however, concerns about the use of limited land resources [10,19], particularly in the context of high commodity prices and a continuously growing population [11,20]. Such concerns may further orient policy makers to invest in the promotion of lignocellulosic biomass, as it can decrease the pressure on prime cropland, if targeted to 'surplus' land [3,5,21]. However, the long cropping cycle of these crops might compete with future food and feed production needs [22]. Knowing which energy crops and where they are likely to be grown is then crucial for a reliable assessment of the biomass supply suitability and of the sustainability of global bioenergy production [23,24]. Indeed, policy estimations frequently assume that enough farmers will choose to grow energy crops if adequately supported with incentives during the start-up phase [25]. This assumption seems, however, to be questioned by a relatively low adoption – approximately 100,000 ha in Europe [7] – compared to their very high technical potential [e.g., 26]. It is therefore important to pursue an up-to-date understanding of farmers' attitudes, behaviours and preferences towards the adoption of perennial energy crops [19,27,28], particularly in the context of farming system innovation [17,29]. Nonetheless, behaviours can vary between farmers and change over time through experience [30–32], eventually becoming harder to predict when facing the choice to plant a perennial species. In fact, this pattern requires researchers to enhance accurate, spatially explicit approaches in order to capture locally-relevant factors, such as soil, climate and logistic factors [33,34,see also 35,36]. However, this enhancement makes a theoretical optimal solution difficult and demanding in terms of computational costs [37].

In brief, despite the clear policy orientation, the economic subsidies and the strong market potential, the actual extent of dedicated bioenergy crops is rather limited [38,39]. Moreover, very few data regarding their actual location are currently available [10]. On the one hand, most of the studies dealing with this type of feedstock tended to assess its production potential based on deterministic approaches and conservative assumptions regarding land use technical potential [5,11]. Examples are constraining the area dedicated to energy crops either (i) to a percentage of the total agricultural area [40,41], (ii) to marginal lands [42–45] or (iii) to the refereeing of "food – feed – nature" or more complex paradigms [22,46–48]. On the other hand, real data shortcomings led research so far to prefer computer simulations to evaluate the potential spatial distribution of energy crops, mostly adapting available process-based modelling [e.g., 49–53]. Nevertheless, where real data are available, empirical models perform better [54] and have even been required to improve the assessment of biomass resource potential at the landscape level [5,14].

To handle real data on energy crop location, recent literature has explored the use of the methods that were originally developed for modelling wild species distribution [e.g., 55,56]. These models commonly use associations between environmental variables and either presence-only or presence-absence data [57]. Presence-only methods have the advantage of relying on very limited datasets, even though they cannot properly handle the role of farming practices in overcoming the environmental constraints to species diffusion [54]. In this work, we used a gradient boosting machine because it is a promising technique used to model species distribution [58]. Also known as boosted regression trees (BRT), this method is an extension of the classification and regression models (also known as CART). A practical advantage of BRT as a tree-based method is that it can handle complex data (i.e., skewed distributions, non-linearity and continuous and categorical data), with no need for variable pre-selection because non-informative predictors are ignored [59].

We restricted our modelling to miscanthus (*Miscanthus x giganteus* Greef et Deuter), which is often considered a promising crop for energy production [7,53,60] and expected to have very high potential yield increase in future decades by breeding for minimal input and improved management [5]. Miscanthus presents high yield potential, requiring low input levels [61–64] and high carbon sequestration capacity [65]; thus, it is likely that it will beneficially reduce greenhouse gas emissions [8]. Furthermore, this crop has advantages over short rotation coppices or other perennial energy crops because it requires very little adaptation of farm equipment [25]. Nevertheless, the effective suitability of the use of miscanthus for energy production depends greatly on the location of this crop and the land use changes that are induced by its adoption [17,66–68].

The aim of our study was to identify realistic prospective locations for miscanthus based on real spatial distribution data. The BRT model of miscanthus spatial location used the crop presence-absence occurrence as a response variable and explanatory variables derived from real farmers' criteria. This approach allowed us to achieve a model that, starting from detailed interviews with miscanthus growers of an existing supply area, was then used to predict miscanthus locations at a regional scale. The results were evaluated under alternative scenarios and distribution constraints.

2. MATERIAL AND METHODS

The general methodology of our approach for modelling and predicting the miscanthus spatial location is illustrated in the graphical abstract. We elaborated a training dataset from the real miscanthus fields that were composing the supply area of the local transformation plant using the farmer's block as the modelling spatial unit. The resulting model was then used to characterise the miscanthus locations and to predict its probable distribution in the study region. Finally, we analysed some probable legislative constraints that were identified in three scenarios.

[FIGURE 1 about here]

2.1. Map of miscanthus presence/absence for the supply area

Miscanthus was established in France only recently, yet it steady increased from the approximately 200 hectares and 87 farmers in 2006 [69] to 2,000 hectares in 2009 [70] and was estimated to occupy a maximum of 3,000 hectares in 2011 [71]. We mapped the real location of miscanthus fields in the supply area of a miscanthus transformation plant – the Bourgogne Pellets cooperative – located in Burgundy, east-central France (Fig. 1). Our focus was on the fields that were planted between 2008 (beginning of the cooperative activities) and 2011. Finally, we covered 386 hectares of miscanthus corresponding to 197 fields managed in total by 75 farmers (Tab. 1).

Then, the real miscanthus fields were associated to the farmer's block as mapped in the French land parcel identification system (LPIS), which is the spatial component of the integrated administration and control system (IACS [72]). We chose the LPIS (reference year 2009, scale 1:5,000) of the French Agency for Service and Payments of the EU Common agricultural Policy subsidies [73] because it provided the highest resolution land use map. It is worth noticing that the spatial relation established in the LPIS between the real agricultural field – a continuous area of land on which a single crop group is cultivated by a single farmer – and the reference parcel – the target for subsidies' payment – is interpreted differently by the Member States [74]. In France, the reference parcel is the "farmer's block", which is defined by the aggregation of neighbouring agricultural fields cultivated by the same farmer (Fig. 2). Each farmer's block is described by the non-localised surface of its land use(s) and a code allowing for the aggregation of the blocks belonging to the same farmland [73]. Of note, miscanthus is not included among the land use classes declared by famers.

The ratio between each real miscanthus field and the related farmer's block was measured and then labelled as "miscanthus presence" the farmer's blocks where miscanthus had a surface greater than 85%. This threshold allowed us to select blocks that can be approximated to miscanthus fields, taking into account the possible geometric mismatch between the real field and the LPIS block. Indeed, in our case study we found that with lower thresholds also mixed farmer's blocks would have been labelled as miscanthus field (cf. Fig. 2b,c), thus introducing a bias in the learning method. Taken together, we obtained 118 farmer's blocks labelled "miscanthus presence" (Tab. 1). The upscaling from field to farm level, to map the miscanthus absence, was realised by mapping the farmland of each farmer who owned at least one farmer's block labelled as "miscanthus presence". This was possible because in the LPIS dataset the farmer's blocks belonging to the same farmland are identified by a unique identity code. We labelled "miscanthus absence" all of the parcels that had none or less than 85% of miscanthus surface. The underpinning hypothesis was to model the farmer's spatial management regarding miscanthus in the context of the overall farm level management to consider his/her main land management units [29]. Finally, the whole dataset was composed of 1939 farmer's blocks.

[FIGURE 2 about here] ; [TABLE 1 about here]

2.2. Explanatory variables composing the training dataset

Martin et al. [75] retrieved a list of the farmers' most relevant criteria through comprehensive interviews. Hereby, we further analysed the results concerning 9 farmers. To date, 7 of these farmers deliver miscanthus to the local transformation plant: their land represents approximately 14% of the total miscanthus surface included in the supply area, which is managed by a total of 75 farmers (Tab. 1). Our focus was on the farmers' criteria at the farmer's block and farm levels for miscanthus that was planted during the 2008-2011 period. We ranked the criteria for their relevance (Tab. 2) according to the frequency in which they occurred in the farmers' decision making processes and then regrouped them as agronomic characteristics, morphological criteria and contextual criteria. The land cover and the inclusion into protected areas were not used in the model, as explained below (section 4.2). The farmers' criteria were then translated to a set of explanatory variables that were used to compose the training dataset (Tab. 2) for the machine learning method. All of the geodata processing was performed in ArcGIS 10 (ESRI; Redlands, CA, USA) with specific tools detailed in the following paragraphs.

2.2.1. Agronomic characteristics

Soil-related properties express the local land suitability and the field accessibility for harvest. The only data available covering the entire area came from the European soil database v.2 (scale scale 1:1,000,000 [76]). The farmer's blocks were intersected with the soil map to retrieve the predominant values for topsoil water capacity and soil texture for each modelling unit. The distance to rivers was used as a proxy of waterlogging – especially in terms of floodability and soil draining capacity – and was calculated with a spatial join between LPIS and BD Carthage® (1:50,000, IGN). Actually, miscanthus performs the better in moist lowland habitats [cf. 77,78] even though exceeding soil water, such in the case of regularly flooded fields, can seriously hamper this rhizomatous crop.

2.2.2. Morphological criteria

The size and shape of a field influence its accessibility to machinery, thus impacting its management. Complex-shaped and/or small fields can be associated with low labour time efficiency [79–81], which is eventually considered disadvantageous for cash crops in farm management. Accordingly, miscanthus was considered by farmers as a relevant alternative because it is a low-intensity labour crop whose work requirement is generally limited to harvesting once the crop is fully established. First, we measured the farmer's block surface. In addition, we evaluated the farmer's block geometry through the classic perimeter/area ratio (proxy of the narrowness) and the shape index (proxy of shape complexity) using Patch Analyst [82]. The shape index was computed by dividing the perimeter by the square root of the polygon area, then adjusting for the circular standard. Hence, it is equal to 1 for polygons close to the shape of a circle, and it increases with increasing shape irregularity. Finally, the local topography was captured as maximum values of elevation and slope for each famer's block, calculated from the BD Alti® (raster resolution of 25 m, IGN) and resumed with Geospatial Modelling Environment [83].

2.2.3. Contextual criteria

Remoteness and accessibility, two complementary features characterising the famer's block within the overall farmland, were approximated as Euclidean distances. We measured how far the farmer's block centroid (computed with XTools v9.1 [84]) was from the transformation plant, the farmland centroid and the three types of roads used by agricultural machinery: *single roadway*, *gravel road* and *pathway* (BD TOPO®, IGN). Notably, the farmland centroid was selected as the best proxy of the farmstead – whose location is unknown due to privacy protection – and was calculated for the multipart feature resulting from the aggregation of all of the parcels sharing the same farmer identity code.

The close proximity of the field boundary to woods can facilitate the presence of wild animals (mainly wild boars) in the cultivated fields, potentially increasing damages to agricultural production [85–87], especially for maize and other cereals [88]. Miscanthus was considered by some farmers as a turnaround to this issue because it is less prone to costly damages than food crops; thus, parcels surrounded by woods are more likely to be targeted for planting miscanthus. Therefore, we measured the boundary that the farmer's block shared with woods as the linear length of the parcel boundaries shared with the neighbouring woods. Only woods larger than 25 ha – located using the Corine Land Cover map year 2006, land cover code 31 [89] – were retained for the analysis. We added a buffer of 30 m to account for shading, for the disruption of machinery circulation due to tree branches and for the consequent reduction of the practicable surface of the farmer's block. In conclusion, we measured the length per farmer's block of the buffered woodland linear boundary using the Geospatial Modelling Environment [83]. Lastly, we defined the closeness to built-up areas as a binary variable (yes/no). The build-up contours were derived from the BD Parcellaire® (IGN) and a buffer of 10 m was added to account for possible geometric errors and nearby roads.

[TABLE 2 about here]

2.3. BRT model set-up and analysis of the results

The presence/absence of miscanthus at the farmer's block level was modelled on a training dataset composed of 1939 farmer's blocks that were characterised using 13 response variables out of the 15 total explanatory variables (Tab.2). We applied BRT that had been implemented for the R statistical environment [90] by the set of functions included in the 'gbm' [91] and 'dismo' [92] packages. The optimal BRT parameterisation was identified by testing different values for the tree complexity (*tc*) and the learning rate (*lr*). The *tc* expresses the interaction depth, where 1 implies an additive model with only a main effect, 2 implies a model with up to 2-way interactions and so forth [58]. The *lr* expresses the contribution of each tree to the growing model. The greater the *tc*, the smaller the *lr* should be kept because it shrinks the contribution of each tree, finally improving the model estimation reliability [93]. The best predictive performances were those that allowed for maximising the area under the receiver operating characteristic (AUC-ROC) that was calculated from a 10-fold cross-validation procedure. Finally, the best trade-off between performances and computation time was achieved with $tc = 3$, $lr = 0.001$ and 5050 trees. The model yielded a miscanthus location probability ranging between 0 and 1 for each farmer's block.

The first goal of our model was to provide an insight into the variables' role to explain the miscanthus location. Although BRT models, likewise other linear combinations of multiple regression trees, are sometimes argued to be less interpretable than simple two-dimensional binary trees [93,94], they can be effectively summarised in different ways. First, they evaluate the role of explanatory variables by ranking their relative influence [91]. The rank derives from the number of times a variable is selected for splitting, weighted by the squared improvement to the model and averaged over all of the trees. Second, partial dependence plots can be obtained to provide a low-dimensional representation of the dependence of the model approximation on the explanatory variables. In fact, these plots show the effect of each predictor on the presence/absence of miscanthus accounting for the average effects of all other variables in the model. Notably, they provide a reliable representation of the effects of each variable, except the case of variables with strong interactions [58].

2.4. Using BRT model to predict miscanthus location in the study region

Understanding the features that could explain the farmers' decision to plant miscanthus in a field is important, but is this knowledge applicable to wider areas? To answer this question, we used the selected best BRT model to predict the miscanthus location probability in the region where the supply area is placed. We ran the model on four out the five departments (NUTS-3 level in the European classification) in the current supply area. The Jura department was excluded because the LPIS data for the year 2009 described only a small portion of the local agricultural area. In the study region (29,017 km^2; 46°10′ to 48°40′N and 3°38′ to 6°49′E) agricultural land covers approximately 17,834 km^2 (Corine Land Cover data [89]), of which 41.2% is managed as arable land and 43.1% as grassland. The remainder consists of permanent crops (1.5%), such as vineyards that produce high quality wine, and heterogeneous areas (14.2%). The great majority of arable lands and grasslands is included also in the LPIS.

2.4.1. Characterising miscanthus predicted location on two thresholds

The characteristics of the farmer's blocks were then compared with two arbitrary thresholds for the predicted miscanthus presence:

(i) 0.1 was chosen according the probability distribution to investigate a possible upper limit for the adoption of miscanthus in the study area, albeit not in greater than 5.24% of the study region agricultural area (cf. section 3.3);

(ii) 0.7, to focus only on the specific (i.e., most probable) miscanthus location.

First the variance homogeneity was assessed for each variable using the Bartlett test and R software. A one-way ANOVA test was performed, and then, the explanatory variable mean values were compared using Tukey's significant difference mean test ($P<0.05$ [95]).

2.4.2. Investigating legislative land use scenarios

We modelled the miscanthus location and predicted its probable location claiming the central role of the farmers' criteria. Nevertheless, the farmers' entrepreneurial choices could be constrained by future evolution both in sectorial policies and regulations. Currently, dedicated energy crops are specifically targeted by environmental regulations to foster sustainability and limit environmental impacts (e.g., Renewable Energy Directive 2009/28/EC) with stricter legislations than those regarding food crops. As mentioned above, bioenergy crop location is an important issue regarding the competition between food, non-food and natural areas at the world scale [e.g., 51]. To address adverse land use change effects that are induced by energy crop expansion, policy makers could consider avoiding the conversion of protected natural areas and of grassland [5,22]. To investigate related possible land use scenarios, we compared three different subsets of the miscanthus predicted location, each representing a different level of potential legislative constraints:

- Business as usual – the unconstrained baseline BRT model where miscanthus is located exclusively depending on the farmer's management criteria.
- Protected areas constraint – provides information on the exclusion of protected areas from the baseline scenario. Hence, we dropped off the farmer's blocks that were included in the most relevant local, regional and national protected areas (Tab. S1 [96]).
- Grassland constraint – accounts for the possible prohibition of replacing grasslands with miscanthus. Such a land use change is debated because it could increase CO_2

emissions and reduce biodiversity [97]. Grassland conversion to other agricultural land is already very limited under European law [98], thus increasing the relevancy of this constraint. Accordingly, for this scenario, we removed the farmer's blocks for which "grassland" was declared as the predominant land use in the LPIS data (Tab. S2) from the baseline scenario results.

Finally, we compared the potential miscanthus area included in the two scenarios and their combination to the baseline scenario. In this way we assessed the effects of high probable land use change constraint on the miscanthus surface of the study area, further detailed for increasing (i.e., 0.1 step) predicted probabilities.

3. RESULTS

The best selected model yielded a value of 0.793 for the AUC-ROC, indicating good predictive performances.

3.1. Important explanatory variables for the supply area

The farmer's block surface is the most important variable for explaining the miscanthus spatial location. In addition, three contextual variables played an important role: the woodland boundary length, the distance to the transformation plant and the distance to the farmland centroid. Altogether, these four variables contributed 73.4% of the model structure (Fig. 3a). The farmer's block elevation showed some influence too, although it was slightly smaller than expected due to chance (i.e., smaller than 7.7%) compared to the remaining variables that were largely above this threshold.

[FIGURE 3 about here]

The partial responses for the presence/absence of miscanthus (Fig. 3b-f) indicate that this crop is more likely to be located in small famer's blocks (but not the smallest ones), with a probability that drastically declines with the increasing of the surface up to 10 hectares (Fig. 3b). In a symmetric way, the probability of miscanthus presence is directly proportional with the increase in length of woodland boundary, although stable for any length greater than approximately 200 meters (Fig. 3c). In addition, the model indicated that miscanthus is more likely to be located immediately around the transformation plant, with a constant increase for any distance greater than 10 km, and a peak at approximately 30 km from it (Fig. 3d). A possible explanation could be that the transformation plant was originally a sugar refinery, thus the surrounding area was more suited for the high demanding sugar beet than for miscanthus. Finally, miscanthus is preferably located, according to the training dataset we used, in land parcels extremely close to the farmland centroid (i.e., less than approximately 200 m) and with an increasing probability within a radius of 2-5 km. In summary, the model indicates that small farmer's blocks with a relatively significant presence of woodland boundary and distance from the farmland centroid are more likely to be considered by farmers for planting miscanthus, especially those blocks located within a radius of 10-30 km from the transformation plant and in plains (elevation smaller than 200 m).

3.2. Characteristics of the predicted miscanthus location in the study region

In the study area, the median surface of a farmer's block is 2.9 hectares and is bigger for arable land (3.8 ha) and smaller for grassland (2.6 ha) and set-aside (0.9 ha) or other land uses (0.4 ha) (see Tab. S2 for details). To evaluate the possible distinctive features of the predicted miscanthus location, we compared the farmer's block properties for two probability thresholds (Tab. 3). The miscanthus presence for the more general threshold (>0.1) was predicted for parcels that were significantly smaller, narrower and had a more complex shape than the remainder of the agricultural area. These parcels are also closer to rivers and have an "easier" morphology (lower slope and altitude), in addition of being farther both from the farmland centroid and from the road. Unexpectedly, the farmer's blocks with a miscanthus location probability greater than 0.1 also had a smaller length of woodland boundary compared to the remaining parcels.

Similarly, the miscanthus presence for the more specific threshold (>0.7) was predicted for parcels smaller and with a more complex shape than the remainders, as well as more distant from rivers and remarkably farther from the farmland centroid and from the road (Tab. 3). No differences emerged instead regarding the narrowness (i.e., perimeter/area ratio) or the slope, whereas the elevation differences were not evaluable using Tukey's test. Noticeably, the miscanthus presence for the higher probability threshold (>0.7) yielded a significantly greater length of woodland boundary. It can be concluded that in our study area, miscanthus would be more likely to be located in somewhat "residual" parcels characterised both by small surfaces and complex shapes that are rather isolated from the farmland centroid and distant from the road although close to the rivers.

Raising the probability threshold – from 0.1 to 0.7, to intercept the more specific pieces of land where farmers might grow miscanthus – reduced the prominence of the morphology but increased the role of the extended woodland boundary. In summary, it seems that the small complex farmer's blocks are weighted for their morphology when considered in general terms for locating miscanthus, whereas the closeness to woodland becomes important when famers might specifically evaluate the miscanthus location. This importance can be due to the greater weight of woodland boundary in reducing the exploitable surface for shadowing, impacting on small land parcels more than the big ones.

[TABLE 3 about here]

3.3. Comparison of the three scenarios

Considering the criteria of the farmers who currently grow miscanthus in the study area, approximately 21% of the farmer's blocks, corresponding roughly to 5% of the total agricultural area, showed a miscanthus location probability greater than 0.1. Only 0.26%, representing 0.06% of the total agricultural area, received a probability greater than 0.7 (Tab. 4). In contrast, the probability that miscanthus might cover a substantial part of the agricultural area (approximately 95%) is quite low (less than 0.1) considering the current criteria of the miscanthus growers.

[TABLE 4 about here] ; [FIGURE 4 about here]

The evaluation of possible legislative scenarios further reduced these results (Fig. 4). A total of 40.3% of the farmer's block surface is included in protected areas (i.e., 604,248 ha) and 51.1% has grassland as the major land use (i.e., 767,387 ha) (Tab. S2). Of note, approximately the 48% of the grassland declared in the LPIS for the study area is in protected areas (i.e., 365,045 ha). Hence, as expected, the exclusion of farmer's blocks in protected areas reduced the total agricultural area by 40%. For the probability thresholds greater than 0.1, the impact was even larger: the potential miscanthus surface was reduced by approximately two-thirds compared to the baseline scenario (Tab. 4). The impact of a possible grassland constraint (i.e., dropping off farmer's blocks with grassland as the major land use) was generally larger than the protected area constraint, with a reduction ranging from 51% and 60% of the baseline scenario.

Noticeably, combining the two constraints and thus avoiding locating miscanthus in protected areas and replacing grassland induced a reduction from 67% to 88% (for increasing probability thresholds), which was much larger than expected. In fact, the farmer's blocks that are currently used for grassland inside protected areas represent only 24.3% of the total agricultural area.

4. DISCUSSION

4.1. The input data and the method

The farmer's block, as mapped in the LPIS, was identified as the spatial modelling unit because it was the best proxy of the real field targeted by farmers to locate miscanthus. This spatially disaggregated agricultural land use map is available, with some differences, all over Europe [74] and supported some recent applications to evaluate the potential for energy crops [34,36]. The main drawback of LPIS, at least in the French and German versions, is that miscanthus is not explicitly recorded. Therefore, additional sources are needed to make the presence-absence modelling of miscanthus (or other bioenergy crops) applicable in different study regions.

Reliable data on (novel) bioenergy crop location and of farmers' criteria that are used to decide their adoption are quite unique, even though they are crucial to assess the accuracy and uncertainty of process-based modelling results for policymakers [10,99]. To date, resource-focused (bottom-up) approaches such as ours have been preferably developed using agent-based methods, which allow for accounting and simulating the farmers' planting decision [10,33] or by using artificial neural networks [35]. We tested the relevance of a novel method, BRT, to provide salient results about real miscanthus location modelling. BRT combines the strengths of decision trees (i.e., delivering a clear support for decision making) and of boosting, which key idea is that the combination of many weak models can provide a better performance than a single strong model because more robust against over-fitting probabilities [58]. Recent applications of BRT models include, for example, investigation on land use changes [100,101] and the spatially explicit assessment of forest harvesting [102] and of forest co-products biomass availability [103].

4.2. Thematic considerations on the findings

Studies investigating the potential of lignocellulosic biomass plantations, especially those based on biophysical potential and economic assessments, may introduce land use constraints (like the "food first") to reduce adverse effects of prospected large-scale biomass cultivation [21,22,49]. However, real-world figures show an uptake that is fairly lower than even more prudent scenarios [7,39,104]. The small-sized farmer's block that was predicted by our model to be relevant for locating miscanthus (Fig. 3b and Tab. 3) seems to provide a possible explanation, at least in our study area. One can presume that the parcels that are adjacent to a farmer's block where miscanthus is likely to be located are equally suitable. However, the farmer's decision criteria – especially those related to the spatial configuration and characteristics of the fields – may drastically reduce the surface that is likely to be grown with this crop (see also [36]). Briefly, during this early stage of the miscanthus adoption, our results indicate that even favourable farmers, who passed the first barrier of the adoption of this new crop, may show their aversion to investing in wide surfaces.

The small and complex-shaped farmer's blocks that were expected to be grown with miscanthus in the study region are also characterised by their great distance from the farm and the roads (Tab. 3). Compared to the general features of the local agricultural area (Tab. 2), these characteristics could provide a different perspective on the definition of "marginal land" thus enhancing the current literature that appears to be mainly focused on the temporary or permanent decline of the productive capacity [5]. Marginal land is frequently defined in an absolute way (e.g., small fields, complex landscape context, inclusion in abandoned areas, etc.), whereas the FAO highlights altogether the presence of "limitations which (…) are severe for sustained application of a given use" [105]. In line with this we deem more relevant to identify the marginal land in a relative way including also the local farmland characteristics, such as the field shape complexity and the distance from the road and from the farmstead (or the collection point). These types of results may complete the research of Harvolk et al. [34], who investigated the ecological potential of miscanthus in marginal lands assuming a random choice of fields. We went further stressing out the attributes of a land parcel that could make it marginal in the farmers' point of view. We extended on this point the considerations by Shortall [106] who analyzed the main definitions of "marginal land", classified either as normative (i.e., "unsuitable for food production" or of "ambiguous low quality") or as predictive (i.e., economical marginality). Whether the former appears to be centered on inherent characteristics of the land evaluated against a specific purpose (mainly food production), the latter makes explicit the possibility that the "marginal" condition might evolve under a different set of price conditions for inputs and the product [106 p. 23]. Our study could add a third point because it deals with the marginality as seen by farmers of a given region linking the field, the farm and the landscape levels. Finally, by tackling together the natural features of the land (agronomic characteristics), its morphological characteristics and the farming contextual aspects (Tab. 2) we addressed the location of miscanthus in marginal lands with a landscape agronomy approach [107].

More in general, in our study area farmers pointed-out the relevant role of current land use in their decision making regarding the field to be planted with miscanthus (Tab. 2). Indeed, the interviews [75] also highlighted that the land use could mask other criteria, thus overlapping with some of the aforementioned explanatory variables. For these reasons, we did not take the land use into account in our model, arguing that its actual role would be expressed by the combination of the other explanatory variables (see Fig. S1 and S2 about the variable interactions). Moreover, farmers claimed interest in the option to plant miscanthus in parcels in protected areas. Miscanthus is actually a low-input crop [64] that could therefore easily meet the protected area rules, yet provide a (greater) income than opting for set-aside or even grassland land use [75]. As the national and European legislation is not yet settled on this matter, we preferred not to consider this variable, as it could express location practices that will be forbidden in the future.

Other variables, such as proximity to built-up, however, can be rather ambivalent in the farmers' decision making regarding miscanthus location. While such a feature raises concerns regarding the possible visual impacts and landscape closing [108,109], some farmers consider proximity to buildings (and settlements) as persuasive because miscanthus is a low/no-input crop, thus conveying a good image of agriculture.

Finally, nothing can be concluded about the preferences for field soil characteristics. Due to a lack of higher resolution data covering the whole area, we used the European soil database that allowed a simplified identification of soil texture and water available content.

4.3. Perspectives for further application

Bioenergy production has complex interactions with other social and environmental systems [1]. In fact, bioenergy policies need to consider regional conditions along with the crop, livestock and forestry sectors [5,22]. However, the impacts and performances of bioenergy production are region- and site-specific, and the effective integration of economic models with a fine-scale land use model still remain a research challenge [23,35].

For example, the distance to the transformation plant has a relevant influence on the adoption of miscanthus because it is a low energy density crop [39]. We addressed this issue by calculating the Euclidian distance from each farmer's block to the plant, even though the real transporting distance should consider the actual road network. Nonetheless, a precise estimation can be difficult because farmers and contractors usually use small local roads (not ever mapped in the available data) and try to avoid crossing villages to prevent nuisances, eventually resulting in non-linear routes. In addition, farmers may use intermediate collection sites (whose location is not easily retrievable) in the farmland, thus splitting the total distance into two or more segments.

A more detailed estimation or a direct survey of distances from the farmer's blocks to the transformation plant, either considering or not considering the intermediate collection sites, could be relevant for improving the actual transportation logistic. This type of model improvement could be used in the predicting step to assess the optimal location of new transformation plants in the study region. Indeed, further scenarios could be developed coupling the predicted miscanthus location probability with an appropriate spatially explicit model to also evaluate the potential yields. However, more work is needed to understand the dynamics between miscanthus supply distribution and the potential location of plants [39,110].

5. CONCLUSION

We proposed a spatially explicit method based on real miscanthus locations to improve the understanding of farmers' criteria and to predict the location of miscanthus for different probability thresholds at a landscape level. Publicly available data were preferred when available to make the model easily replicable. Altogether, the main strength and novelty of the model and the prediction we proposed are to stick with such complex reality from the farmers' perspectives with a very fine-scale resolution, finally spanning from the field to the landscape level. This proposition is advantageous because it allows for to grasp all of the complexity of the farmers' styles while avoiding the flattening required by some modelling approaches on few farmers' types (to avoid complex models and restrain the working hypotheses). More accurate modelling approaches would require shifting to case-based reasoning methods [111], which are in the early phase of development concerning the treatment of spatially explicit problems [112–114]. In contrast, the validity domain of our work could be somewhat dependent on the characteristics of the study region. Therefore, we look forward to replicating the model in different contexts (e.g., in terms of regional topography and field pattern structure) to better understand its sensitivity to the study region characteristics.

Our results provide a snapshot of a static economic context, namely characterised by low prices for miscanthus, which can be considered as a baseline potential. Alternative scenarios could address variations in the list and the weight of location decision criteria or foster higher miscanthus adoption to meet policy expectations. Nevertheless, we maintain that the direct involvement of farmers is required to ensure that the model properly grasps the complexity of the local farming systems and provides reliable salient results for policy making.

ACKNOWLEDGEMENTS

We are grateful to Philippe Béjot (Bourgogne Pellets Cooperative) who kindly provided the fundamental data to map the real miscanthus fields. We also warmly thank the farmers involved in this project for the time they spent explaining their work. Special thanks to the Agence de Service et de Paiement for granting access to the LPIS data and also to Amandine Durpoix and Jean-Marie Trommenschlager (INRA SAD-ASTER) for their support in the LPIS processing. The Institut National de l'Information Géographique et Forestière, the Muséum national d'Histoire naturelle (via the National inventory of natural heritage) and the European Commission (Joint Research Centre) provided the geographic datasets, the protected areas and the soil maps for this research, which was partly funded by the FUTUROL project and by the French state innovation agency OSEO. This work has been funded also under the EU seventh Framework Programme by the LogistEC project N° 311858: Logistics for Energy Crops' Biomass. The views expressed in this work are the sole responsibility of the authors and do not necessary reflect the views of the European Commission.

REFERENCES

[1] IPCC. Summary for Policymakers. Spec. Rep. Intergov. Panel Clim. Change, Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press; 2011, p. 3–26.

[2] EREC. European Renewable Energy Council – RE-thinking 2050: a 100% renewable energy vision for the European Union. 2010.

[3] Rahman MM, B. Mostafiz S, Paatero JV, Lahdelma R. Extension of energy crops on surplus agricultural lands: A potentially viable option in developing countries while fossil fuel reserves are diminishing. Renew Sustain Energy Rev 2014;29:108–19.

[4] Krewitt W, Nienhaus K, Kleßmann C, Capone C, Stricker E, Graus W, et al. Role and potential of renewable energy and energy efficiency for global energy supply. 2009.

[5] Chum H, Faaij APC, Moreira JR, Berndes G, Dhamija P, Dong H, et al. Bioenergy. IPCC Spec. Rep. Renew. Energy Sources Clim. Change Mitig., Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press; 2011, p. 209–331.

[6] Monforti F, Bódis K, Scarlat N, Dallemand J-F. The possible contribution of agricultural crop residues to renewable energy targets in Europe: A spatially explicit study. Renew Sustain Energy Rev 2013;19:666–77.

[7] Don A, Osborne B, Hastings A, Skiba U, Carter MS, Drewer J, et al. Land-use change to bioenergy production in Europe: implications for the greenhouse gas balance and soil carbon. GCB Bioenergy 2012;4:372–91.

[8] Bessou C, Ferchaud F, Gabrielle B, Mary B. Biofuels, greenhouse gases and climate change. A review. Agron Sustain Dev 2011;31:1–79.

[9] Bentsen NS, Felby C. Biomass for energy in the European Union-a review of bioenergy resource assessments. Biotechnol Biofuels 2012;5:1–10.

[10] Li R, di Virgilio N, Guan Q, Feng S, Richter GM. Reviewing models of land availability and dynamics for biofuel crops in the United States and the European Union. Biofuels Bioprod Biorefining 2013;7:666–84.

[11] Dornburg V, Vuuren D van, Ven G van de, Langeveld H, Meeusen M, Banse M, et al. Bioenergy revisited: Key factors in global potentials of bioenergy. Energy Environ Sci 2010;3:258–67.

[12] Smeets EMW, Lewandowski IM, Faaij APC. The economical and environmental performance of miscanthus and switchgrass production and supply chains in a European setting. Renew Sustain Energy Rev 2009;13:1230–45.

[13] Asbjornsen H, Hernandez-Santana V, Liebman M, Bayala J, Chen J, Helmers M, et al. Targeting perennial vegetation in agricultural landscapes for enhancing ecosystem services. Renew Agric Food Syst 2013:1–25.

[14] Heaton EA, Schulte LA, Berti M, Langeveld H, Zegada-Lizarazu W, Parrish D, et al. Managing a second-generation crop portfolio through sustainable intensification: Examples from the USA and the EU. Biofuels Bioprod Biorefining 2013;7:702–14.

[15] Holzmueller EJ, Jose S. Biomass production for biofuels using agroforestry: potential for the North Central Region of the United States. Agrofor Syst 2012;85:305–14.

[16] Howard DC, Burgess PJ, Butler SJ, Carver SJ, Cockerill T, Coleby AM, et al. Energyscapes: Linking the energy system and ecosystem services in real landscapes. Biomass Bioenergy 2012;55:17–26.

[17] Zegada-Lizarazu W, Elbersen HW, Cosentino SL, Zatta A, Alexopoulou E, Monti A. Agronomic aspects of future energy crops in Europe. Biofuels Bioprod Biorefining 2010;4:674–91.

[18] Cosentino SL, Patanè C, Sanzone E, Copani V, Foti S. Effects of soil water content and nitrogen supply on the productivity of Miscanthus × giganteus Greef et Deu. in a Mediterranean environment. Ind Crops Prod 2007;25:75–88.

[19] Ostwald M, Jonsson A, Wibeck V, Asplund T. Mapping energy crop cultivation and identifying motivational factors among Swedish farmers. Biomass Bioenergy 2013;50:25–34.

[20] Acevedo MF. Interdisciplinary progress in food production, food security and environment research. Environ Conserv 2011;38:151–71.

[21] Dauber J, Brown C, Fernando AL, Finnan J, Krasuska E, Ponitka J, et al. Bioenergy from “surplus” land: environmental and socio-economic implications. BIORISK – Biodivers Ecosyst Risk Assess 2012;7:5–50.

[22] Lovett AA, Sunnenberg GM, Richter GM, Dailey AG, Riche AB, Karp A. Land Use Implications of Increased Biomass Production Identified by GIS-Based Suitability and Yield Mapping for Miscanthus in England. Bioenergy Res 2009;2:17–28.

[23] Nassar AM, Harfuch L, Bachion LC, Moreira MR. Biofuels and land-use changes: searching for the top model. Interface Focus 2011;1:224–32.

[24] Sanscartier D, Deen B, Dias G, MacLean HL, Dadfar H, McDonald I, et al. Implications of land class and environmental factors on life cycle GHG emissions of Miscanthus as a bioenergy feedstock. GCB Bioenergy 2013:n/a–n/a. doi: 10.1111/gcbb.12062

[25] Sherrington C, Bartley J, Moran D. Farm-level constraints on the domestic supply of perennial energy crops in the UK. Energy Policy 2008;36:2504–12.

[26] EEA. European Environment Agency - How much bioenergy can Europe produce without harming the environment? Copenhagen, Denmark: European Environment Agency; 2006.

[27] Glithero NJ, Wilson P, Ramsden SJ. Prospects for arable farm uptake of Short Rotation Coppice willow and miscanthus in England. Appl Energy 2013;107:209–218.

[28] Sherrington C, Moran D. Modelling farmer uptake of perennial energy crops in the UK. Energy Policy 2010;38:3567–78.

[29] Rizzo D, Marraccini E, Lardon S, Rapey H, Debolini M, Benoît M, et al. Farming systems designing landscapes: land management units at the interface between agronomy and geography. Geogr Tidsskr-Dan J Geogr 2013;113:71–86.

[30] Primdahl J. Agricultural landscapes as places of production and for living in owner’s versus producer’s decision making and the implications for planning. Landsc Urban Plan 1999;46:143–50.

[31] Farmar-Bowers Q, Lane R. Understanding farmers’ strategic decision-making processes and the implications for biodiversity conservation policy. J Environ Manage 2009;90:1135–44.

[32] Guillem EE, Barnes AP, Rounsevell MDA, Renwick A. Refining perception-based farmer typologies with the analysis of past census data. J Environ Manage 2012;110:226–35.

[33] Alexander P, Moran D, Rounsevell MDA, Smith P. Modelling the perennial energy crop market: the role of spatial diffusion. J R Soc Interface 2013;10:20130656.

[34] Harvolk S, Kornatz P, Otte A, Simmering D. Using existing landscape data to assess the ecological potential of Miscanthus cultivation in a marginal landscape. GCB Bioenergy 2013:15 p.

[35] Li R, Guan Q, Merchant J. A geospatial modeling framework for assessing biofuels-related land-use and land-cover change. Agric Ecosyst Environ 2012;161:17–26.

[36] Moser D, Eckerstorfer M, Pascher K, Essl F, Zulka KP. Potential of genetically modified oilseed rape for biofuels in Austria: Land use patterns and coexistence constraints could decrease domestic feedstock production. Biomass Bioenergy 2013;50:35–44.

[37] Dunnett AJ, Adjiman CS, Shah N. A spatially explicit whole-system model of the lignocellulosic bioethanol supply chain: an assessment of decentralised processing potential. Biotechnol Biofuels 2008;1:1–17.

[38] Christou M, Alexopoulou E, Panoutsou C, Monti A. Overview of the markets for energy crops in EU27. Biofuels Bioprod Biorefining 2010;4:605–19.

[39] Alexander P, Moran D, Smith P, Hastings A, Wang S, Sünnenberg G, et al. Estimating UK perennial energy crop supply using farm-scale models with spatially disaggregated data. GCB Bioenergy 2013:14 p.

[40] Ericsson K, Nilsson LJ. Assessment of the potential biomass supply in Europe using a resource-focused approach. Biomass Bioenergy 2006;30:1–15.

[41] Callesen I, Grohnheit PE, Ostergard H. Optimization of bioenergy yield from cultivated land in Denmark. Biomass Bioenergy 2010;34:1348–62.

[42] Smeets EMW, Faaij APC, Lewandowski IM, Turkenburg WC. A bottom-up assessment and review of global bio-energy potentials to 2050. Prog Energy Combust Sci 2007;33:56–106.

[43] Fiorese G, Guariso G. A GIS-based approach to evaluate biomass potential from energy crops at regional scale. Environ Model Softw 2010;25:702–11.

[44] Hellmann F, Verburg PH. Spatially explicit modelling of biofuel crops in Europe. Biomass Bioenergy 2011;35:2411–24.

[45] Tenerelli A, Carver S. Multi-criteria, multi-objective and uncertainty analysis for agro-energy spatial modelling. Appl Geogr 2012;32:724–36.

[46] De Wit M, Faaij A. European biomass resource potential and costs. Biomass Bioenergy 2010;34:188–202.

[47] Fischer G, Prieler S, van Velthuizen H, Berndes G, Faaij A, Londo M, et al. Biofuel production potentials in Europe: Sustainable use of cultivated land and pastures, Part II: Land use scenarios. Biomass Bioenergy 2010;34:173–87.

[48] Haughton AJ, Bond AJ, Lovett AA, Dockerty T, Sünnenberg G, Clark SJ, et al. A novel, integrated approach to assessing social, economic and environmental implications of changing rural land-use: a case study of perennial biomass crops. J Appl Ecol 2009;46:315–22.

[49] Beringer T, Lucht W, Schaphoff S. Bioenergy production potential of global biomass plantations under environmental and agricultural constraints. GCB Bioenergy 2011;3:299–312.

[50] Dufossé K, Gabrielle B, Drouet J-L, Bessou C. Using Agroecosystem Modeling to Improve the Estimates of N2O Emissions in the Life-Cycle Assessment of Biofuels. Waste Biomass Valorization 2013;4:593–606.

[51] Pogson M, Hastings A, Smith P. How does bioenergy compare with other land-based renewable energy sources globally? GCB Bioenergy 2013;5:513–24.

[52] Xu X, Li S, Fu Y, Zhuang D. An analysis of the geographic distribution of energy crops and their potential for bioenergy production. Biomass Bioenergy 2013;59:325–35.

[53] Hastings A, Clifton-Brown J, Wattenbach M, Mitchell CP, Stampfl P, Smith P. Future energy potential of Miscanthus in Europe. GCB Bioenergy 2009;1:180–96.

[54] Estes LD, Bradley BA, Beukes H, Hole DG, Lau M, Oppenheimer MG, et al. Comparing mechanistic and empirical model projections of crop suitability and productivity: implications for ecological forecasting. Glob Ecol Biogeogr 2013;22:1007–18.

[55] Evans JM, Fletcher RJ, Alavalapati J. Using species distribution models to identify suitable areas for biofuel feedstock production. GCB Bioenergy 2010;2:63–78.

[56] Trabucco A, Achten WMJ, Bowe C, Aerts R, Orshoven JV, Norgrove L, et al. Global mapping of Jatropha curcas yield based on response of fitness to present and future climate. GCB Bioenergy 2010;2:139–51.

[57] Pearson RG. Species' distribution modeling for conservation educators and practitioners. LinC 3, 2007, 54-89.

[58] Elith J, Leathwick JR, Hastie T. A working guide to boosted regression trees. J Anim Ecol 2008;77:802–13.

[59] De'ath G. Boosted trees for ecological modeling and prediction. Ecology 2007;88:243–51.

[60] Anderson E, Arundale R, Maughan M, Oladeinde A, Wycislo A, Voigt T. Growth and agronomy of Miscanthus × giganteus for biomass production. Biofuels 2011;2:167–83.

[61] Ercoli L, Mariotti M, Masoni A, Bonari E. Effect of irrigation and nitrogen fertilization on biomass yield and efficiency of energy use in crop production of Miscanthus. Field Crops Res 1999;63:3–11.

[62] Lewandowski I, Scurlock JMO, Lindvall E, Christou M. The development and current status of perennial rhizomatous grasses as energy crops in the US and Europe. Biomass Bioenergy 2003;25:335–61.

[63] Zub HW, Brancourt-Hulmel M. Agronomic and physiological performances of different species of Miscanthus, a major energy crop. A review. Agron Sustain Dev 2010;30:201–14.

[64] Cadoux S, Riche AB, Yates NE, Machet J-M. Nutrient requirements of Miscanthus x giganteus: Conclusions from a review of published studies. Biomass Bioenergy 2012;38:14–22.

[65] Kahle P, Beuch S, Boelcke B, Leinweber P, Schulten H-R. Cropping of Miscanthus in Central Europe: biomass production and influence on nutrients and soil organic matter. Eur J Agron 2001;15:171–84.

[66] Clifton-Brown JC, Stampfl PF, Jones MB. Miscanthus biomass production for energy in Europe and its potential contribution to decreasing fossil fuel carbon emissions. Glob Change Biol 2004;10:509–18.

[67] Hastings A, Clifton-Brown J, Wattenbach M, Mitchell CP, Smith P. The development of MISCANFOR, a new Miscanthus crop growth model: towards more robust yield predictions under different climatic and soil conditions. GCB Bioenergy 2009;1:154–70.

[68] Hillier J, Whittaker C, Dailey G, Aylott M, Casella E, Richter GM, et al. Greenhouse gas emissions from four bioenergy crops in England and Wales: Integrating spatial estimates of yield and soil carbon balance in life cycle analyses. GCB Bioenergy 2009;1:267–81.

[69] Gurtler J-L, Féménias A, Blondy J. Agriculture Énergie 2030: fiche-variable production de bioénergies 2009.

[70] Association France Miscanthus. Cultivons l'énergie de demain avec le Miscanthus ! 2009. http://www.cgb-france.fr/IMG/pdf/Brochure_miscanthus_2009.pdf last accessed February 22

[71] AEBIOM. European Bioenergy Outlook 2013. Bruxelles, Belgium: European Biomass Association; 2013.

[72] Inan HI, Sagris V, Devos W, Milenov P, van Oosterom P, Zevenbergen J. Data model for the collaboration between land administration systems and agricultural land parcel identification systems. J Environ Manage 2010;91:2440–54.

[73] ASP. Agence de Service et de Paiement [Agency for Service and Payment]. Registre parcellaire graphique anonyme [French Anonymous Land Parcel Identification System] 2009.

[74] Sagris V. Land Parcel Identification System conceptual model: development of geoinfo community conceptual model. PhD thesis. University of Tartu (Estonia), 2013.

[75] Martin L, Wohlfahrt J, Le Ber F, Benoît M. Perennial biomass crops allocation: a French case study regarding miscanthus. Espace Géogr 41, 2012, 133-147.

[76] Panagos P, Van Liedekerke M, Jones A, Montanarella L. European Soil Data Centre: Response to European policy support and public data requirements. Land Use Policy 2012;29:329–38.

[77] Barney JN, Mann JJ, Kyser GB, DiTomaso JM. Assessing habitat susceptibility and resistance to invasion by the bioenergy crops switchgrass and Miscanthus × giganteus in California. Biomass Bioenergy 2012;40:143–54.

[78] Maughan M, Bollero G, Lee DK, Darmody R, Bonos S, Cortese L, et al. Miscanthus × giganteus productivity: the effects of management in different environments. GCB Bioenergy 2012;4:253–65.

[79] Amiama C, Bueno J, Álvarez CJ. Influence of the physical parameters of fields and of crop yield on the effective field capacity of a self-propelled forage harvester. Biosyst Eng 2008;100:198–205.

[80] Gónzalez XP, Marey MF, Álvarez CJ. Evaluation of productive rural land patterns with joint regard to the size, shape and dispersion of plots. Agric Syst 2007;92:52–62.

[81] Herrmann C, Prochnow A, Heiermann M. Influence of chopping length on capacities, labour time requirement and costs in the harvest and ensiling chain of maize. Biosyst Eng 2011;110:310–20.

[82] Rempel R s., Kaukinen D, Carr AP. Patch Analyst and Patch Grid. Centre for Northern Forest Ecosystem Research, Thunder Bay, Ontario: Ontario Ministry of Natural Resources; 2012.

[83] Beyer HL. Geospatial Modelling Environment. 2012.

[84] Data East Soft L. XTools Pro - Extension for ArcGIS. data East Soft, LLC; 2012.

[85] Amici A, Serrani F, Rossi CM, Primi R. Increase in crop damage caused by wild boar (Sus scrofa L.): the "refuge effect." Agron Sustain Dev 2012;32:683–92.

[86] Calenge C, Maillard D, Fournier P, Fouque C. Efficiency of spreading maize in the garrigues to reduce wild boar (Sus scrofa) damage to Mediterranean vineyards. Eur J Wildl Res 2004;50:112–20.

[87] Hofman-Kamińska E, Kowalczyk R. Farm Crops Depredation by European Bison (Bison bonasus) in the Vicinity of Forest Habitats in Northeastern Poland. Environ Manage 2012;50:530–41.

[88] Herrero J, García-Serrano A, Couto S, Ortuño VM, García-González R. Diet of wild boar Sus scrofa L. and crop damage in an intensive agroecosystem. Eur J Wildl Res 2006;52:245–50.

[89] EEA. European Environment Agency. Corine Land Cover 2006 seamless vector data 2012.

[90] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing; 2013.

[91] Ridgeway G. gbm: Generalized Boosted Regression Models. 2013.

[92] Hijmans RJ, Phillips S, Leathwick J, Elith J. dismo: Species distribution modeling. 2013.

[93] Hastie T, Tibshirani R, Friedman JH. The elements of statistical learning: data mining, inference, and prediction. Springer; 2008.

[94] Witten IH, Frank E, Hall MA. Data mining: practical machine learning tools and techniques. Burlington, MA: Morgan Kaufmann; 2011.

[95] Hothorn T, Bretz F, Westfall P, Heiberger RM, Schuetzenmeister A. multcomp: Simultaneous Inference in General Parametric Models. 2013.

[96] Muséum national d'Histoire naturelle. National inventory of natural heritage 2013.

[97] Rettenmaier N, Köppen S, Gärtner SO, Reinhardt GA. Life cycle assessment of selected future energy crops for Europe. Biofuels Bioprod Biorefining 2010;4:620–36.

[98] Isselstein J, Jeangros B, Pavlu V. Agronomic aspects of biodiversity targeted management of temperate grasslands in Europe–a review. Agron Res 2005;3:139–51.

[99] Augustenborg CA, Finnan J, McBennett L, Connolly V, Priegnitz U, Müller C. Farmers' perspectives for the development of a bioenergy industry in Ireland. GCB Bioenergy 2012;4:597–610.

[100] Müller D, Leitão PJ, Sikor T. Comparing the determinants of cropland abandonment in Albania and Romania using boosted regression trees. Agric Syst 2013;117:66–77.

[101] Petty JT, Strager MP, Merriam EM, Ziemkiewicz PF. Scenario analysis and the watershed futures planner: predicting future aquatic conditions in an intensively mined Appalachian watershed. In Craynon JR, editor. Environmental considerations in energy product ion, Englewood: Society for Myning, Metallurgy, and Exploration, Inc.; 2013, p.5-19.

[102] Levers C, Verkerk PJ, Müller D, Verburg PH, Butsic V, Leitão PJ, et al. Drivers of forest harvesting intensity patterns in Europe. For Ecol Manag 2014;315:160–72.

[103] Idir K. The spatially explicit assessment of forest biomass availability at regional level. Proceeding 8th Conf. Sustain. Dev. Energy Water Environ. Syst., Dubrovnik, Croatia: Faculty of Mechanical Engineering and Naval Architecture; 2013.

[104] Hastings A, Clifton-Brown J, Wattenbach M, Stampfl P, Mitchell CP, Smith P. Potential of Miscanthus grasses to provide energy and hence reduce greenhouse gas emissions. Agron Sustain Dev 2008;28:465–72.

[105] CGIAR. Chapter 2 - definitions and context. Rep. Study CGIAR Res. Priorities Marg. Lands, 1997.

[106] Shortall OK. "Marginal land" for energy crops: Exploring definitions and embedded assumptions. Energy Policy 2013;62:19–27.

[107] Benoît M, Rizzo D, Marraccini E, Moonen AC, Galli M, Lardon S, et al. Landscape agronomy: a new field for addressing agricultural landscape dynamics. Landsc Ecol 2012;10:1385–94.

[108] Dockerty T, Appleton K, Lovett A. Public opinion on energy crops in the landscape: considerations for the expansion of renewable energy from biomass. J Environ Plan Manag 2012;55:1134–58.

[109] Pointereau P, Bochu J-L, Couturier C, Coulon F, Arnal A, Giorgis S. Les impacts environnementaux et paysagers des nouvelles productions énergétiques sur les parcelles et bâtiments agricoles. Ministère de l'Agriculture et de la Pêche; 2009.

[110] Thomas A, Bond A, Hiscock K. A GIS based assessment of bioenergy potential in England within existing energy systems. Biomass Bioenergy 2013;55:107–21.

[111] Martin L, Ber FL, Wohlfahrt J, Bocquého G, Benoît M. Modelling farmers' choice of miscanthus allocation in farmland: a case-based reasoning model. Proc. 2012 Int. Congr. Environ. Model. Softw., 2012.

[112] Du Y, Liang F, Sun Y. Integrating spatial relations into case-based reasoning to solve geographic problems. Knowl-Based Syst 2012;33:111–23.

[113] Holt A, Benwell GL. Applying case-based reasoning techniques in GIS. Int J Geogr Inf Sci 1999;13:9–25.

[114] Osty PL, Le Ber F, Lieber J. Raisonnement à partir de cas et agronomie des territoires. Rev Anthropol Connaiss 2008;2:169–93.

APPENDIX A. SUPPLEMENTARY DATA

Table S1. Characteristics of the protected areas considered for the scenarios.

Table S2. Characteristics of the farmer's blocks according to the major land uses in the study area.

Figure S1. Diagram of the most important pairwise interactions for the complete (all the variables listed in Tab. 4) and the partial models (the variables used for the final model).

Figure S2 Three-dimensional partial dependence plots for the strongest pairwise interactions in the selected model.

FIGURE CAPTIONS

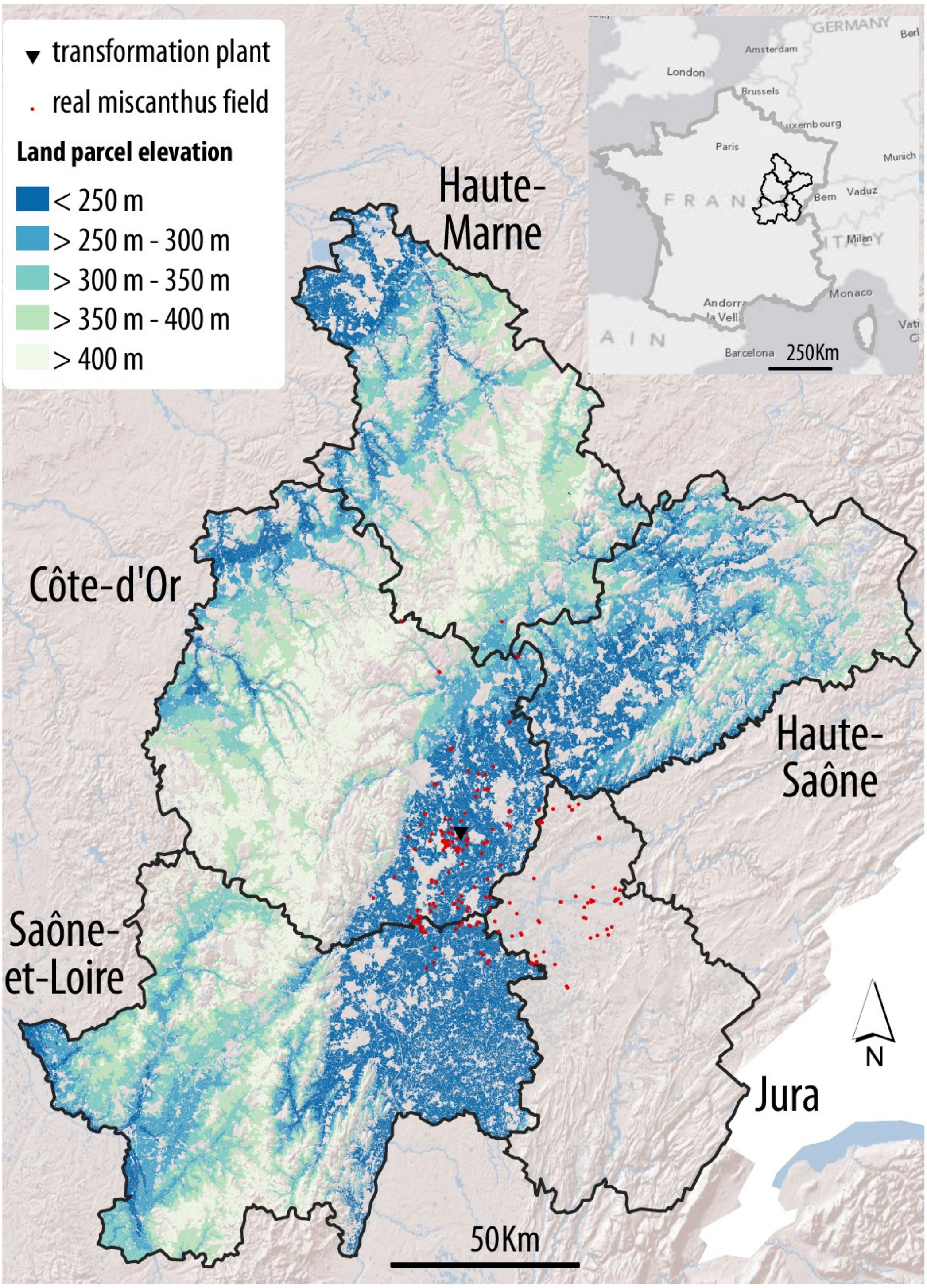


Fig. 1 Location of the supply area within the study region and topography of the agricultural area (source: land parcel identification system, year 2009 and IGN data).

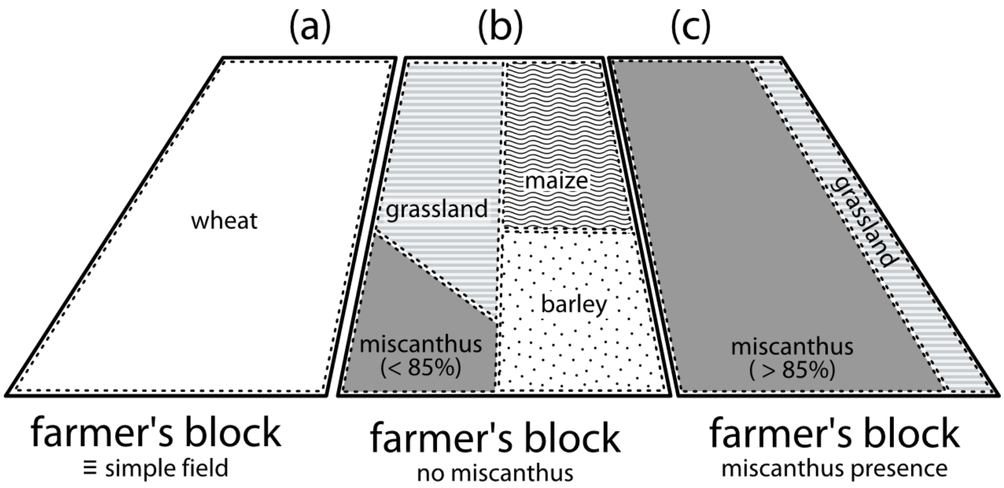


Fig. 2 Diagram comparing the agricultural field and the farmer's block as mapped in the French land parcel identification system. Fig. 2a: simple field (i.e., one land use) coincident with a farmer's block. Fig. 2b: farmer's block composed of several fields (i.e., different land uses) one of which is miscanthus extending for less the 85% (i.e., under the threshold of "miscanthus presence"). Fig. 2c: farmer's block composed of several fields one of which is miscanthus extending for more than 85% of the surface (i.e., above the threshold of "miscanthus presence").

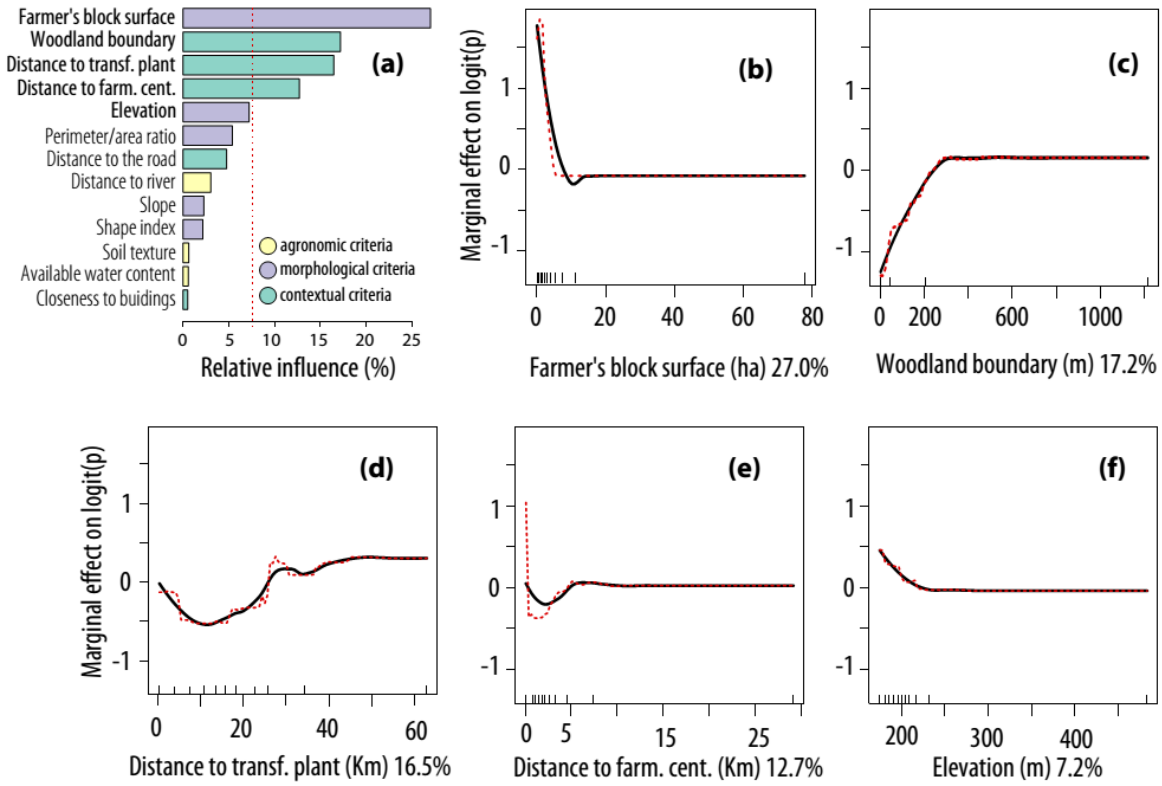


Fig. 3 Main results of the miscanthus location model for the supply area. Fig. 3a: relative importance of miscanthus location explanatory variables; values are in percentage, normalised to sum to 100 and longer bars represent greater relative influence of the explanatory variable. The red dotted line marks the threshold beyond which the relative influence is greater than expected to chance. Fig. 3b-f: marginal effects of the first five explanatory variables on the probability (expressed as logit(p)) of presence-absence of miscanthus. The partial dependence plots illustrate the change in the logit of the probability (log-odds on the y-axis) along a given explanatory variable (x-asis), holding all other constant: higher median values correspond to a higher likelihood of famer's block selection for locating miscanthus. Percentage values express the variable relative importance for the overall model. Solid black lines show the smoothed fitted function, dashed red lines show the original value. Rug plots at the inside bottom of plots show distribution of parcels across that variable, in deciles.

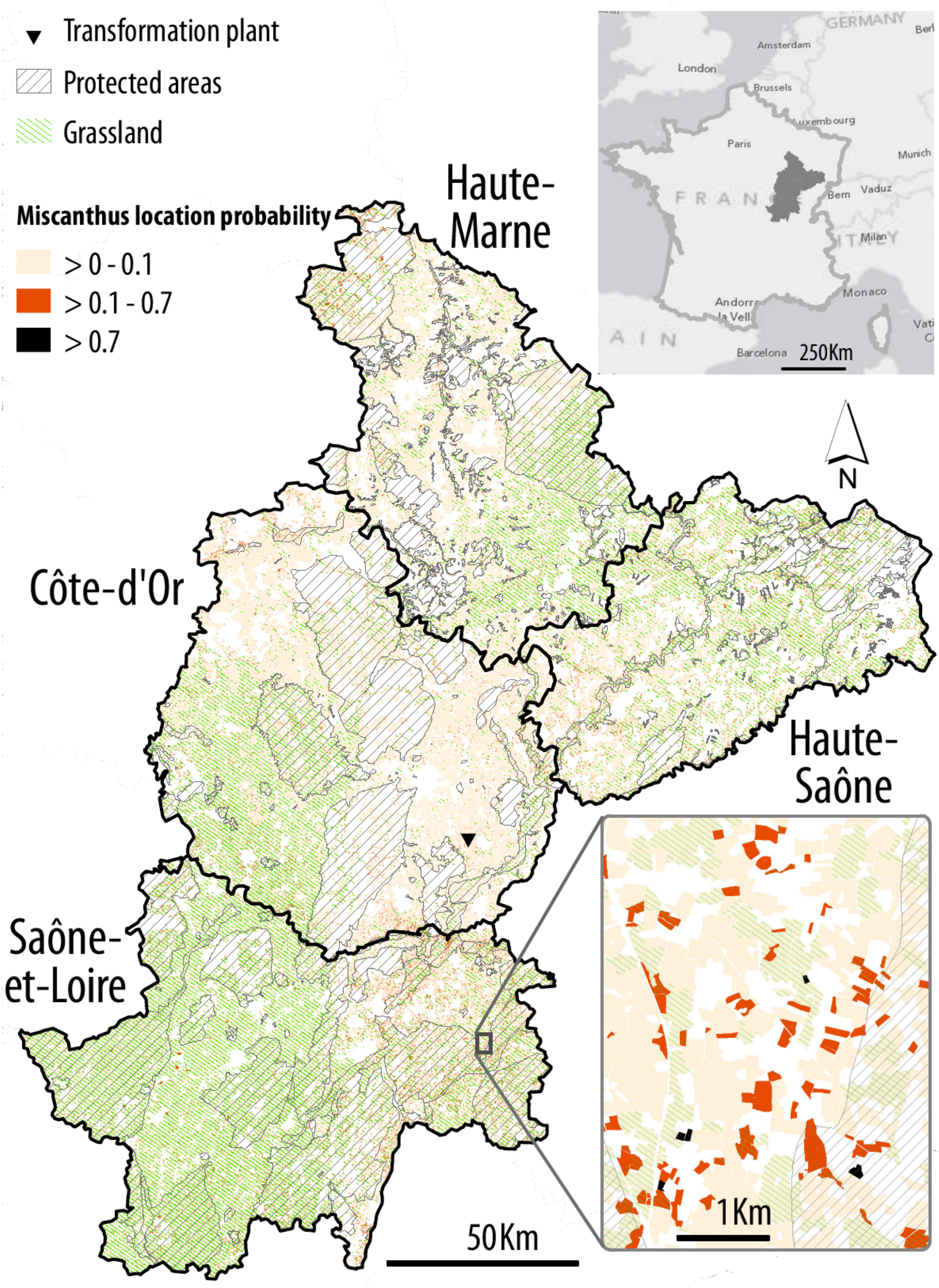


Fig. 4 Probability of miscanthus spatial location predicted with the BRT model for the study region and constraints used in the prospective scenarios.

Miscanthus spatial location as seen by farmers: a machine learning approach to model real criteria

Rizzo, Martin, Wohlfahrt – INRA SAD-ASTER,

corresponding author: ridavide@gmail.com (Rizzo Davide)

TABLES

Table 1 Distribution of the real miscanthus data for the 2008-2011 period. Source: statistics on real miscanthus field map.

		2008	2009	2010	2011	Total
Surface of real fields (ha)	Total	3.5	100.7	204.3	76.9	385.4
	Range	1.0-2.5	0.2-4.0	0.3-14.3	0.2-15.2	0.2-15.2
	Mean (s.d.)	-	1.5 (0.9)	2.3 (2.1)	2.0 (2.7)	1.96 (1.95)
Number of new farmers/year		2	35	28	10	75
Number of fields		2	69	88	38	197
Number of farmer's blocks[1] (miscanthus "presence" only)		1	46	48	23	118

1 Farmer's blocks in the land parcel identification system correspondent to the real fields

Table 2 Explanatory variables adapted from farmers' criteria described by Martin et al. [75], and response variables (N=13) used to model miscanthus spatial location; (*) indicates categorical variables. For each variable essential statistics allow to compare the training dataset and the study area dataset used to model miscanthus location.

Explanatory variables *(Farmers' decision criteria)*	Relevance	Response variables (N=13)	Description for the farmer's block	values	Learning dataset (N = 1939)			Study area (N=263630)		
					range	mean (s.d.)	median	range	mean (s.d.)	median
Agronomic characteristics										
Soil water availability	••••	AWC*	Available water content in the topsoil	medium (100-140 mm/m), high (140 -190 mm/m), very high (>190 mm/m), missing data	-	-	high	-	-	high
Waterlogging	••	RivDist	Distance to river as proxy of floodability and/or draining soils	meters	0-3245	456 (466)	333	0-7009	550.62 (719.38)	324
Soil mechanical properties	•	Text*	Soil texture	coarse (1) to fine (4)	-	-	2	-	-	3
Morphological criteria										
Size	••••	PHa	Surface	hectares	0.02-77.90	5 (6.2)	2.98	0-383.43	5.69 (8.46)	2.86
Geometry	••••	PSI	Shape complexity adjusted for circular standard (1 if close to the circle shape, greater otherwise	-	1.01-4.47	1.51 (0.4)	1.39	1-9.41	1.5 (0.39)	1.39
		PPAR	Perimeter/area ratio (the greater is the value, the narrower is the farmer's block)	meters/hectares	54.3-3599.7	412.7 (355.7)	306.20	31-16669.2	415.15 (369.94)	310.10
Slope	•	PSlope	Maximum slope	percentage	0.0-43.1	3.6 (3.5)	2.20	0-111.5	10.7 (8.3)	8.30
Topography	•	PAlt	Maximum elevation	meters	175-438	206 (28)	200	115-1215	312 (99)	301
Contextual criteria										
Remoteness	•••	PlantDist	Distance from the farmer's block centroid to transformation plant	meters	479-62729	17899 (11825)	15803	390-169308	75522 (33993)	74037

Explanatory variables ***(Farmers' decision criteria)***	**Relevance**	**Response variables (N=13)**	**Description for the farmer's block**	**values**	**Learning dataset (N = 1939)**			**Study area (N=263630)**		
					range	**mean (s.d.)**	**median**	**range**	**mean (s.d.)**	**median**
		FarmDist	Distance from the farmer's block centroid to farmland centroid (see text for detail)	meters	68-29140	3245 (3228)	2165	0-191615	3015 (4562)	1732
Accessibility	•••	RoadDist	Distance to the closest road	meters	0-221	5.4 (17.7)	2	0-584	7.1 (25.2)	2
Proximity to wood	•••	PWbm	Linear length of the farmer's block boundaries shared with the neighbour wood (with a 30m buffer)	meters	0-1217	53 (141)	0	0-5863	93 (242)	0
Proximity to build-up	••	PUrb*	Contact with built-up area (with a 10m buffer)	yes, no	-	-	no	-	-	no
Other criteria										
Land cover	•••	-	Prevalent land use type (e.g., grassland or arable)	Cf. table S2						
Inclusion in protected areas	••	-	Whether the land block is inside in a local or regional level protected area	Cf. table S1						
Response variable										
Miscanthus presence-absence		MisPoP*	"Presence" if miscanthus surface >85% of the farmer's block. "Absence" for block belonging to a farmland with >1 miscanthus presence	0 (absence), 1 (presence)	-	-	no			

Table 3 Comparison of the explanatory variable mean values. Two thresholds of miscanthus location probability are considered: general (threshold 0.1) and specific (threshold 0.7). For each probability threshold, values in the same row with different letters (a, b) are significantly different (P<0.05)

Miscanthus location probability	General		Specific	
	≤ 0.1	> 0.1	≤ 0.7	> 0.7
Number of farmer's blocks	absence	presence	absence	presence
	207520	56110	262968	662
Farmer's block surface (ha)	6.8 [a]	1.4 [b]	5.7 [a]	1.4 [b]
Perimeter/area ratio (m/ha)	402.3 [a]	462.8 [b]	415.1 [a]	417.5 [a]
Shape index (unitless)	1.49 [a]	1.55 [b]	1.50 [a]	1.57 [b]
Woodland boundary (m)	100 [a]	70 [b]	93 [a]	203 [b]
Distance to river (m)	572 [a]	472 [b]	551 [a]	356 [b]
Slope (%)	10.9 [a]	9.9 [b]	10.7 [a]	11.1 [a]
Elevation (m)	316 [a]	294 [b]	312 [*]	272 [*]
Distance to the farmland centroid (m)	2913 [a]	3391 [b]	3003 [a]	7628 [b]
Distance to the road (m)	5 [a]	14 [b]	7 [a]	28 [b]

* test non-applicable due to variance heterogeneity

Table 4 Proportion of agricultural area predicted for planting miscanthus. Decreasing location probability (columns) and the three different scenarios are illustrated. ( ) Grey columns highlight the lower limit of the two thresholds identified for comparing location probability (see Tab. 3 and the text for details)

	Miscanthus location probability	>0.9	>0.8-0.9	>0.7-0.8	>0.6-0.7	>0.5-0.6	>0.4-0.5	>0.3-0.4	>0.2-0.3	>0.1-0.2	0-0.1
Baseline scenario	Number of farmer's blocks	43	202	417	759	1472	2667	5066	10945	34539	207520
	% on total farmer's blocks	0.02	0.08	0.16	0.29	0.56	1.00	1.92	4.15	13.10	78.72
	Area (ha)	56.1	2879.8	570.4	1,049.1	1,956.9	3,412.1	6,583.8	14,338.9	50,433.6	1,422,147.0
	% on total agricultural area	0.004	0.02	0.04	0.07	0.13	0.23	0.44	0.95	3.36	94.76
	(Cumulated)	(0.004)	(0.02)	(0.06)	(0.13)	(0.26)	(0.49)	(0.93)	(1.88)	(5.24)	(100.00)
Outside protected areas	Area (ha)	17.4	117.4	299.6	553.8	920.9	1640.3	3183.0	7178.3	27115.7	855552.6
	% of baseline reduction	*69*	*58*	*47*	*47*	*53*	*52*	*52*	*50*	*46*	*40*
Land use different from grassland	Area (ha)	22.2	122.5	258.3	448.1	824.6	1491.7	2908.0	6150.9	20565.1	700649.0
	% of baseline reduction	*60*	*56*	*55*	*57*	*58*	*56*	*56*	*57*	*59*	*51*
Outside protected areas with a land use different from grassland	Area (ha)	6.7	53.3	151.6	275.1	444.4	790.7	1536.4	3386.8	12021.5	475570.5
	% of baseline reduction	*88*	*81*	*73*	*74*	*77*	*77*	*77*	*76*	*76*	*67*

Miscanthus spatial location as seen by farmers: a machine learning approach to model real criteria

Rizzo, Martin, Wohlfahrt – INRA SAD-ASTER (France)

corresponding author: ridavide@gmail.com (Rizzo Davide)

APPENDIX A. SUPPLEMENTARY DATA

Table S1 Characteristics of the protected areas considered for the scenarios. Source: Muséum national d'Histoire naturelle, 2013: National inventory of natural heritage.

	Name	**Definition**
Included in the scenario	"Arrêté de protection des biotopes"	Departmental-level protected area defined by-law to protect local habitat or endangered species
	"Réserves biologiques"	Natural reserve fostering the protection of woodlands and related habitats
	"Conservatoire des espaces naturels"	Local associations, managing natural site also through land consolidation and knowledge dissemination
	National Park	
	"Réserves naturelles régionales"	Regionally defined long term protection of outstanding natural areas
	RAMSAR sites	Site protected under the Convention on Wetlands of International Importance, especially as Waterfowl Habitat
	Biosphere reserves	Part of the Man and Biosphere Reserve Programme
	IBA (in French: ZICO)	Important Bird Area, recognized as being globally important habitat for the conservation of bird populations
	ZNIEFF 1 et 2	Natural Zone of Ecological interest for flora and fauna
	SCI (in French: SIC)	Natura 2000 - Site of Community Importance as defined in the European Commission Habitats Directive (92/43/EEC)
	SPA (in French: ZPS)	Natura 2000 - Special Protection Area designated under the European Union Directive on the Conservation of Wild Birds (79/409/CEE)
Not included in the scenario	Zone de conservation chasse et faune sauvage	Protected areas for hunting activities and wildlife conservation
	Réserves maritimes	Marine reserves
	Conservatoire du littoral	Public organisation created in 1975 to ensure the protection of outstanding natural areas on the coast, banks of lakes and stretches of water of 10 square kilometres or more
	Regional Park	

Table S2 Characteristics of the farmer's blocks according to the major land uses in the study area. Values are expressed in hectares. Source: statistics and adaptation from the land parcel identification system, year 2009*.

	composition	Mean size (s.d.)	Median size	surface (inside or on the edge of parks)	ratio on total
Arable N= 104128	Wheat, maize, barley and other winter cereals; rapeseed, sunflower and other oilseed and industrial crops, fibre plants; rice, grain legumes, fodder crops	6.8 (9.5)	3.8	707929 (32%)	47.2%
Set-aside N=8557	Set-aside, either with no production, for industrial crop only or of other type	1.9 (3.3)	0.9	15901 (41%)	1.1%
Grassland N=142688	Permanent and temporary pastures, summer pastures and moorland	5.4 (7.9)	2.6	767387 (48%)	51.1%
Other N=8257	Seed crops, fruit groves and orchards, vineyards, nut-trees, olive groves, vegetables and flowers, sugarcane, other crops	1.2 (2.7)	0.4	9610 (51%)	0.6%
Total N=263630	total Land Parcel Identification System (LPIS) area	5.7 (8.5)	2.9	1500827 (40%)	-

*ASP (2009) Agence de Service et de Paiement [Agency for Service and Payment]. Registre parcellaire graphique anonyme [French Anonymous Land Parcel Identification System].

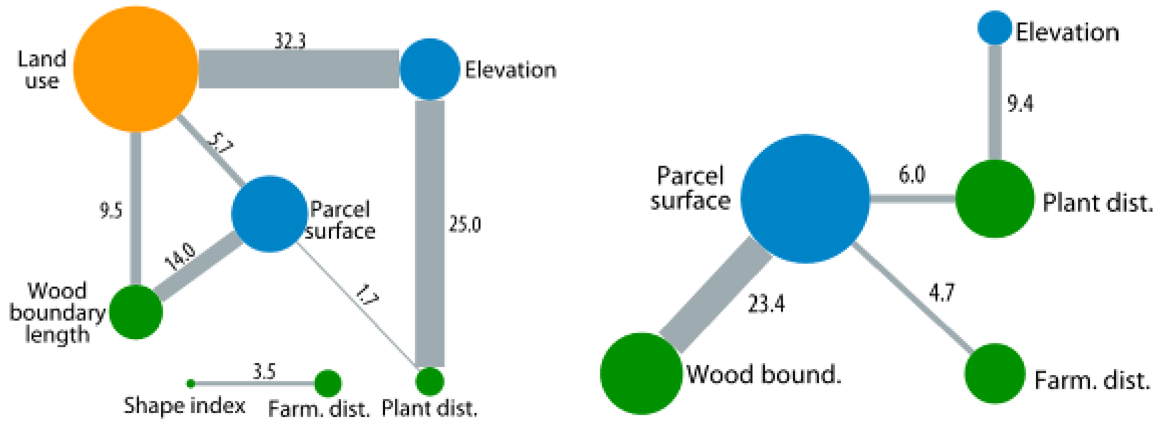


Figure S1 Diagram of the most important pairwise interactions for the complete (all the variables listed in Tab. 4: see article for details) and the partial models (the variables used for the final model). Circle diameter is proportional to the variable relative importance. Connector width is proportional to the relative strength of the interaction, whose value is indicated by the number on the connection.

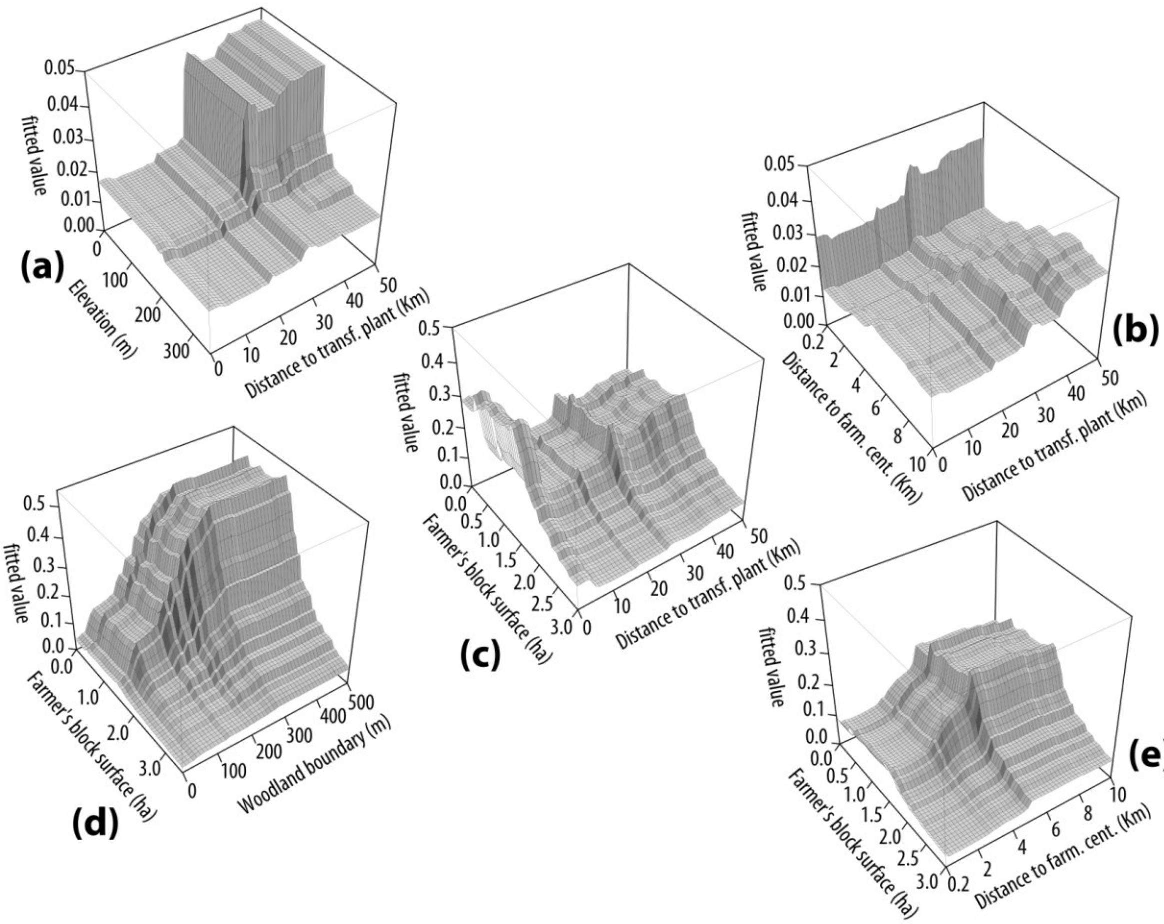


Figure S2 Three-dimensional partial dependence plots for the strongest pairwise interactions in the selected model.