UNITE MIXTE DE RECHERCHE
EN ECONOMIE PUBLIQUE

JOINT RESEARCH UNIT IN
PUBLIC ECONOMICS

2016
CAHIERS DE RECHERCHE
WORKING PAPERS
2016/01

Land use predictions on a regular grid at different scales and with easily accessible covariates

R. Chakir; T. Laurent; A. Ruiz-Gazen; C. Thomas-Agnan; C. Vignes

UMR Economie Publique
Avenue Lucien Brétignières – 78850 Grignon
16 rue Cl. Bernard – 75005 Paris
Tel. +33 (0)1 30 81 53 30
Fax. +33 (0)1 30 81 53 68
http://www.grignon.inra.fr/economie-publique

Land use predictions on a regular grid at different scales and with easily accessible covariates

Raja Chakir[*], Thibault Laurent[**], Anne Ruiz-Gazen[***], Christine Thomas-Agnan[***], and Céline Vignes[**]

[*]UMR Economie Publique INRA-AgroParisTech, BP01, 78850 Thiverval-Grignon, France
[**]Toulouse School of Economics (CNRS/GREMAQ), 21 allée de Brienne, 31042 Toulouse, France
[***]Toulouse School of Economics (GREMAQ), 21 allée de Brienne, 31042 Toulouse, France

March 2, 2015

1 Abstract

We investigate in this paper the extent to which a simple model based on easily accessible covariates can lead to predictions of land use comparable to the ones obtained from a more complete but costly information. Land use is measured by physical occupation such as defined in the Teruti-Lucas survey and we consider five categories: urban, farming, forests, pastures and natural land. Our approach consist in two steps: the first one is a model of land use at the level of the Teruti-Lucas points whereas the second one derives a land use prediction on a regular mesh at several scales. The first step yields fine scale predictions and the second one aggregates them on the mesh cells. We study the quality of the predictions as a function of the size of the mesh cells. We show that with easily accessible variables, we obtain an acceptable quality for the Teruti-Lucas points and that the quality improves noticeably as early as the first aggregation step.

2 Introduction

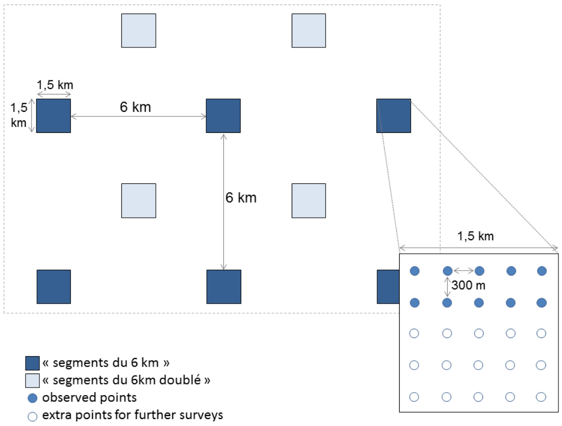
Obtaining fine level data on land use in France such as the Teruti-Lucas 2010 survey of the French Ministry of agriculture can be difficult and costly. We investigate the quality of alternative solutions consisting in using a prediction model based on easily accessible variables. We consider predicting land use on regular meshes at different scales. The problems are first of all the choice of the model at the level of the Teruti-Lucas points, the choice of an aggregation rule to go from the Teruti-Lucas points level to the mesh cell level and finally the choice of a criterion for assessing the quality at different scales. In section 3, we present the data as well as the models (multinomial logits and classification trees) and the Brier score method for assessing prediction quality. In section 4, we report the results of model fitting and, in section 5, the quality of predictions as a function of the aggregation method and the geographical scale.

3 Data and methods

Our region of interest is the Midi-Pyrénées region which is the largest region in France (8.3 % of the whole territory and 3020 municipalities on January 2013). It is a rural region which contains only 4.5 % of the metropolitan population (2011).

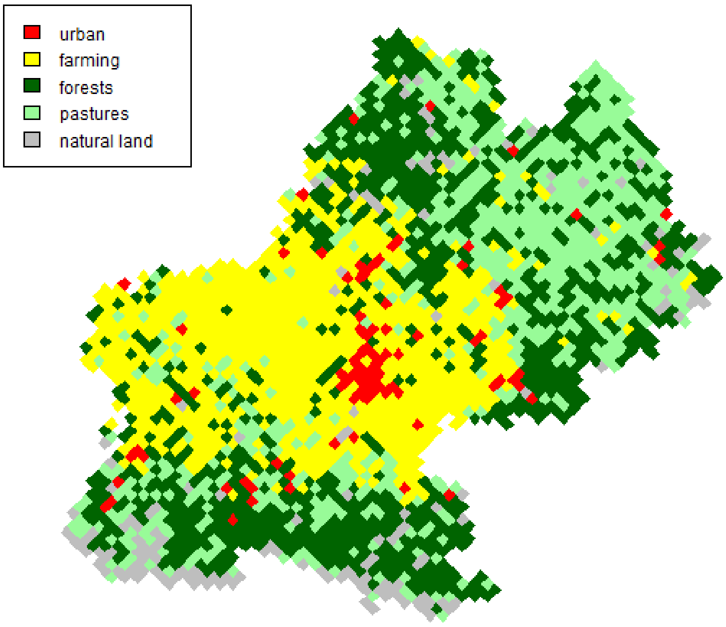
Since 2005, the Teruti-Lucas survey contains information about the evolution of land use on segments containing 25 points out of which only 10 of them are presently used to collect the data (see Figure 1). Our sample thus contains 25317 points for the Midi-Pyrénées region each year. The segments are grouped in two subsets which are named in French "segment du 6km" and "segment du 6km doublé."

Figure 1: Teruti-Lucas survey: segments and points



We focus on physical occupation of the land as opposed to the functional occupation (socio-economic destination) which is also collected. Coding of the physical occupation is done with two nested structures, one with 57 levels and one with 122 levels. We consider the 57 levels structure and group the levels in the following five categories: urban, farming, forests, pastures, natural land. Figure 2 presents the main land use by segment in the Region.

Figure 2: Main land use by segment



The explanatory variables we consider are coming from diverse data bases at several different scales. Table 1 describes the source of each explanatory variable. More details about the data sets and variables can be found in Chakir et al. [2015]. The finest information comes from the Corine Land Cover data and the altitude (from IGN). Using normalized neighborhood matrices from spatial econometrics (distance threshold of 1400 m), we have computed spatially lagged variables (mean of neighboring values) for continuous variables. For the Corine Land Cover variable which is categorical, we have used the main use in each segment as lagged variable.

Table 1: Data sources

name	geographical level	source	year	unit
land use	6km segment	Teruti-Lucas	2010	-
CLC	zones (>25 ha)	Corine Land Cover	2006	-
altitude	grid (250m)	BDAlti de l'IGN	-	meters
soil constitution	UCS zones	BGSF (GISSOL)	1998	-
main surface				-
base material				-
evolution of soil texture				-
presence of a waterproof layer				-
meteorology	grid 25x25km	Agri4cast	2010	-
annual minimum of daily temperature				degrees C
annual maximum of daily temperature				degrees C
annual mean of daily temperature				degrees C
annual sum of rain quantity				millimeters
mean speed of wind				km/h
land and empty meadow price meadow (more than 70 ha)	32 NRA	Agreste	2010	actual € /ha
socio-economic data	municipalities	Insee	2010	-
population density				inhabitants/km^2
percentage of farmers				%
percentage of executives				%
metropolitan center				-

In order to evaluate the predictive ability of our models, we divide the sample in a learning sample for building the model and a test sample for evaluating the prediction quality: for this we use the natural division described above into "segment du 6km" (testing subsample, 12660 points) and "segment du 6km doublé" (learning subsample, 12657 points) so that their spatial distribution be regular.

Table 2: Frequencies of land use categories in 2010 for the two samples (number of Teruti-Lucas points and %)

landuse	label	learning sample count	learning sample %	testing sample count	testing sample %	total count	total %
1	urban	933	7.4	909	7.2	1 842	7.3
2	farming	3 344	26.4	3 252	25.7	6 596	26.1
3	forests	3 906	30.9	4 051	32.0	7 957	31.4
4	pastures	3 231	25.5	3 279	25.9	6 510	25.7
5	natural land	1 246	9.8	1 155	9.1	2 401	9.5
Total		12 660	100	12 646	100	25 306	100

The models we use in the first step of our procedure are the multinomial logit models [McFadden, 1974] and the classification trees (CART, Breiman et al. [1984]). We adopt two point of views: a statistical one which aims at selecting explanatory variables that best predicts land use and an econometrics one that selects variables suggested by economic theory. After fitting a multinomial model, we obtain estimated probabilities for each land use category and each point. If we want to go further and get predictions, we need a rule for constructing predictions from the estimated probabilities. A classical one is to predict that land use corresponding to the maximal probability (over all uses) but we could also do a random draw from the estimated multinomial. This last method is quite bad at the level of points but will turn out to be competitive at higher aggregation levels. Reversely, for classification trees, one directly gets predictions for each point since each leaf has an associated land use. But if an estimate of the probabilities is needed, one can use the empirical frequency corresponding to the group of leaves associated with each given land use category. For both methods, once the predictions computed, it is easy to compute the rates of correct predictions for each category.

As we mentioned earlier, we consider regular square meshes at several scales. The finest level, denoted by A_0, corresponds to the Teruti-Lucas points, and the coarsest one, denoted by A_7, corresponds to the whole Midi-Pyrénées region. Mesh A_1 is constructed so as to contain a unique Teruti-Lucas segment and to form a tiling of the territory: its squares are centered on the barycenter of the 10 points of the Teruti-Lucas segment and will be called unit squares hereafter. Their sides have a length of 4.2 kilometers and this mesh comprises 2579 such squares (we see those on Figure 2). We then construct several successive aggregations of these unit squares until level A_6 where each square contains 1024 unit squares. This is summarized in Table 3. To

Table 3: Characteristics of the meshes

Mesh	Number of aggregated "unit squares"	Approximate area	Number of points per square	Total number of squares
A_1	1	18 km^2	1 à 10	2 579 squares
A_2	4	72 km^2	1 à 40	689 squares
A_3	16	288 km^2	4 à 160	192 squares
A_4	64	1 152 km^2	10 à 640	59 squares
A_5	256	4 608 km^2	184 à 2 559	20 squares
A_6	1 024	18 432 km^2	184 à 6 605	8 squares

aggregate the estimated probabilities at level A_k say, we just average the estimated probabilities of the cells of the previous aggregation level A_{k-1}. However to define predictions at level A_k, we have the choice between two solutions. The first one is similar to what we did at the level of the points and consists in predicting the category which has maximal estimated probability. The second one is to draw a random sample from the multinomial with the estimated probabilities at the point level and then derive the empirical frequencies of each use at the mesh level. It is important to note that it is not possible anymore to evaluate the quality of prediction at the mesh level since we have no true value at that level to compare with. For this reason, we propose to use the weighted Brier score which consists in a small modification of the well known Brier score (see Brier [1950]) taking into account the number of aggregated points in each mesh cell. In Chakir et al. [2015], we prove that the weighted Brier score is always decreasing when aggregating a given level to a coarser one. In what follows we consider different models for land use prediction, first at the level of the Teruti-Lucas point and then, on regular grids at different scales. We also illustrate the use of the weighted Brier score to help the decision maker in choosing the aggregation level.

4 Prediction at the Teruti-Lucas point level

The models are fitted and the predictions are made at the Teruti-Lucas points level. The statistical methods we compare are the multinomial logit regression (MNL) as detailed in McFadden [1974] and the classification tree method using the CART algorithm as detailed in Breiman et al. [1984]. The first fitted model is a multinomial logit model called "Econometrics" and denoted by MNL-E where the choice of the explanatory variables is motivated by the economics literature. According to the empirical economics literature [Lubowski, 2002] on land use, the variables that have an impact on the land use decision are the revenues associated to the land use or some proxy variables like input and output prices, the public aid, the density of population (as a proxy of the revenue of urban land use) and soil and climate variables such as altitude, slope, soil chemistry, temperature and rain fall. We also refine the model by taking into account statistical arguments. In parallel, we consider a classification tree using the whole set of explanatory variables and trying to maximize the correctly classified rate. Finally, taking into account both the MNL and the classification tree approaches we estimate six different models and compare them (see Table 4).

Table 4: Correctly classified rates

Variables in the model	MNL	trees
CLC2 + altitude	MNL-S : 65,15 %	TREE-S : 64.92 %
CLC2 + altitude + MAT + TEXT + landprice + cadre	-	TREE-Ct : 64.83 %
CLC2 + altitude + mintemp + meanwind + maxtemp + altitude_lag	MNL-Cm : 65.37 %	-
CLC2 + altitude+landprice+density + mintemp + maxtemp + meantemp + rain + TEXT	MNL-E : 65.12 %	TREE-E : 65.04 %

Table 5: Model MNL-E coefficients (Significance levels: ***: 0.001 **: 0.01 * : 0.05)

variable	farming use	forest	pastures	natural land
constant	-4.089	-1.831	-5.693**	-2.808
CLC2-12	-1.267	-1.144	-1.788	0.506
CLC2-13	0.909	0.762	0.123	2.257***
CLC2-14	-18.33	0.142	-20.700	-2.143
CLC2-21	4.359***	2.223***	3.271***	1.505***
CLC2-22	3.882***	2.050***	2.479***	1.672***
CLC2-23	2.871***	3.096***	4.124***	2.063***
CLC2-24	3.155***	2.562***	3.362***	1.47***
CLC2-31	2.300***	5.459***	2.827***	2.822***
CLC2-32	1.989***	3.488***	3.605***	4.018***
CLC2-33	-13.350	1.820	1.292	3.912***
CLC2-41	2.104	20.966	19.498	20.055
CLC2-51	-15.465	3.969***	1.968	5.175***
altitude	0.0003	0.002***	0.003***	0.002***
landprice	0.0001**	-0.0002***	0.0000	-0.0003***
density	-0.0003	-0.0004	-0.001**	-0.001*
mintemp	0.178***	0.064	-0.095*	0.041
maxtemp	0.161**	0.099*	0.071	0.162**
meantemp	-0.266*	-0.215*	-0.023	-0.247*
rain	-0.001	-0.001*	-0.001	-0.002***
TEXT-1	-0.08	0.497	-0.092	-0.246
TEXT-2	0.129	0.183	-0.233	-0.368
TEXT-3	0.413	0.406	0.100	0.178
TEXT-4	0.285	0.204	-0.075	-0.174

The results from model MNL-E are presented in table 5 and indicate that the categories of the variable CLC2 are significant. The price of the farming lands has a significant and positive impact on the farming, forests and pastures compared to the urban use. The population density has no significant impact on the farming and forests use but has a negative and significant impact on the pastures and natural land use. A result that may be surprising is the nonsignificance of the categories of the soil constitution. A possible explanation might be that the information contained in these variables is already taken into account by the variables CLC2 and altitude. Concerning the quality of model MNL-E in terms of prediction, the correctly classified points rate is 65,12 % which is quite comparable with the models MNL-Cm and MNL-S (see table 9).

Table 6: Model MNL-Cm coefficients (Significance levels: ***: 0.001 **: 0.01 * : 0.05)

variable	farming use	forest	pastures	natural land
constant	-10.667***	-3.228	-7.607***	-7.356**
CLC2-12	-1.261	-1.173	-1.814	0.479
CLC2-13	0.743	0.702	0.036	2.122***
CLC2-14	-18.128	0.304	-20.643	-1.535
CLC2-21	4.461***	2.24***	3.295***	1.625***
CLC2-22	3.989***	2.152***	2.46***	1.899***
CLC2-23	2.938***	3.13***	4.125***	2.072***
CLC2-24	3.209***	2.62***	3.368***	1.542***
CLC2-31	2.357***	5.528***	2.821***	2.867***
CLC2-32	2.067***	3.604***	3.634***	4.177***
CLC2-33	-12.12	1.884	1.474	4.269***
CLC2-41	2.535	20.978	19.6	20.19
CLC2-51	-15.422	4.006***	2.001	5.262***
altitude	0.005*	0.004**	0.006***	0.006***
mintemp	0.067*	0.045	-0.09**	0.009
meanwind	0.084***	-0.023	0.006	0.001
maxtemp	0.209***	0.028	0.094*	0.115*
altitude_lag	-0.005*	-0.002	-0.004*	-0.003

Table 7: Model MNL-S coefficients (Significance levels: ***: 0.001 **: 0.01 * : 0.05)

variable	farming use	forest	pastures	natural land
constant	-1.814***	-2.688***	-3.035***	-2.813***
CLC2-12	-1.259	-1.123	-1.782	0.542
CLC2-13	0.767	0.79	0.092	2.222***
CLC2-14	-18.040	0.495	-20.440	-1.335
CLC2-21	4.513***	2.307***	3.323***	1.718***
CLC2-22	3.967***	2.246***	2.495***	2.024***
CLC2-23	2.848***	3.141***	4.221***	2.089***
CLC2-24	3.165***	2.643***	3.464***	1.586***
CLC2-31	2.271***	5.549***	2.892***	2.890***
CLC2-32	1.945***	3.661***	3.681***	4.225***
CLC2-33	-11.510	2.235*	1.535	4.560***
CLC2-41	2.612	21.006	19.870	20.264
CLC2-51	-15.582	3.989***	2.085*	5.252***
altitude	-0.001***	0.002***	0.003***	0.002***

When searching to optimize the correctly classified prediction rate using a stepwise method on the learning sample, we obtain a subset of explanatory variables which differs from the ones in model MNL-E. The model incorporating these variables is called MNL-Cm (see table 6). In particular, the socio-economics variables (landprice and density) and the soil texture variable are not selected. The variables meantemp and rain disappear while meanwind and altitude_lag are incorporated in the model. The six selected variable are CLC2, altitude, mintemp, wind speed, maxtemp and altitude_lag. The correctly classified rate is only improved by a small amount with a value of 65.37 %.

Concerning the classification trees, we use the CART algorithm and all the available explanatory variables except those in the neighborhood[1]. If we use the default input parameters in the R

[1] Note that if we introduce the neighborhood variables, the altitude_lag variable is selected while altitude is

function `rpart`, the complexity criterion is set to 0.01 and the tree is so much pruned that it takes into account only two explanatory variables (CLC2 and altitude). In order to get a number of explanatory variables of the same order as MNL-Cm, we set the initial complexity criterion to 0.0001. To obtain the tree TREE-Ct, we use the function `rpart` which provides a maximal tree and the function `prune` which gives a pruned tree following the procedure proposed by Breiman et al. [1984]. Using this function leads to a complexity criterion value equal to 0.001. The resulting tree is very complex. It contains 16 nodes and takes into account six variables (CLC2, altitude, landprice, the percentage of executives and, concerning the soil constitution, the base material and the texture). The meteorological variables are not selected while the socio-economic variables and the soil constitution variables are selected. Given its complexity, this tree will not be detailed further. The correctly classified rate is equal to 64.83 %, which is slightly smaller than the rate for MNL-E and MNL-Cm.

From TREE-Ct, we construct a new and simpler tree called TREE-S which maximizes the correctly classified rate by pruning successively the TREE-Ct (see table 8). The best correctly classified rate obtained is 64.92 % and is associated with TREE-S which contains ten nodes and only two variables: CLC2 and altitude (see figure 3).

Table 8: Evolution of the correctly classified rate as a function of the complexity criterion of the tree

cp	number of nodes	correctly classified rate	
0.001	16	64.83 %	(TREE-Ct)
0.002	14	64.90 %	
0.0025	13	64.87 %	
0.003	11	64.92 %	
0.005	10	64.92 %	(TREE-S)
0.007	8	64.48 %	
0.008	7	64.15 %	
0.01	6	63.47 %	

We also construct a tree based on the explanatory variables of the econometric MNL model and obtain the tree called TREE-E that is made of 15 nodes and selects 5 variables (with no meteorological variable). Its correctly classified rate is equal to 65.04 % and is similar to the rate for the MNL-E.

Finally, if we construct a MNL model with the explanatory variables CLC2 and altitude selected by TREE-S we obtain a correctly classified rate equal to 65.15 % similar to the one for the MNL-E and only slightly smaller that one of the MNL-Cm.

Note that the main difference between models MNL-E and MNL-S is the nonsignificance of the altitude variable corresponding to the farming use for the econometric model while this coefficient differs significantly from zero at the level 1/1000 in the simple model.

Table 9 provides different quality indicators for the three proposed multinomial logit models. We can see that the three models lead to very similar results.

With a correct prediction of around two thirds of the points the results are already quite good at the point level. However, we do not recover the variability of the land use we observe at the

not. Moreover, the percentage of executives in the neighborhood is also selected but without any improvement on the correctly classified rate which becomes 64.73 %.

Figure 3: Tree TREE-S

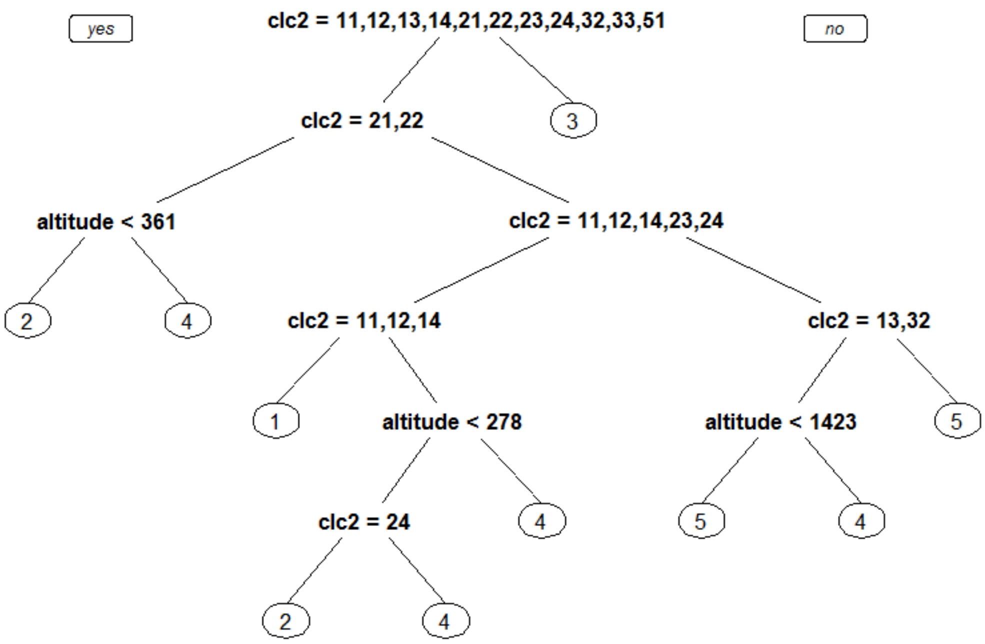


Table 9: Comparison of the quality of multinomial logistic regression models

Indicator	MNL-E	MNL-Cm	MNL-S
BIC	26 255.1	26 197.3	26 197.1
AIC	25 540.3	25 661.2	25 780.2
Mc Fadden R^2	0.3251	0.3206	0.3166
correctly classified rate	65.12 %	65.37 %	65.15 %

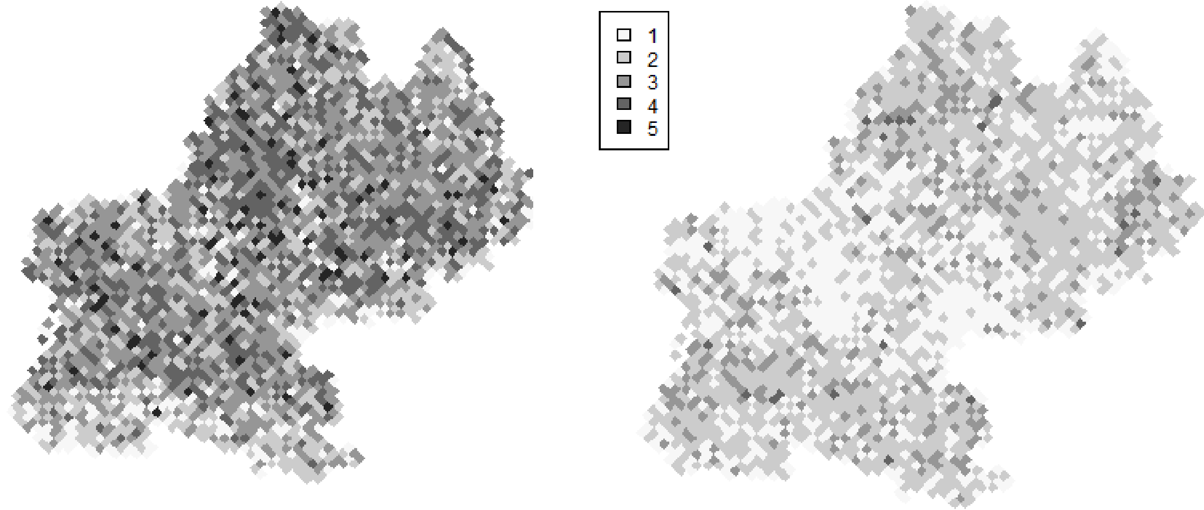
segment level using our predictions. While for instance, 67 % of the segments have at least three different observed land use, this proportion becomes 14.7 % for the predicted land use. Figure 4 illustrates also that point.
At this stage we may wonder whether we can improve on the prediction quality by looking at aggregated spatial levels.

5 Prediction on grids at different scales

The final objective of a user is to obtain prediction and estimated probabilities of land use on a regular grid that he can use either to plot some maps or as an ingredient in a model. One of the potential questions may be the choice of the grid scale. Moreover, in order to get prediction on a regular grid, the estimated probabilities obtained at the point level have to be aggregated. Some proposal for the previous two questions are to be found in what follows.

Concerning the aggregation method, the standard methodology in this situation, whether for MNL or trees, is to calculate first the estimated probability mean of the points that belong to the same mesh cell. Given this aggregated estimated probability, we can predict the land use of the cell as the one associated with the maximal estimated probability. Another possibility is to predict by drawing at random according to a multinomial distribution with the aggregated

Figure 4: Number of land use per segment, observed land use (left panel) or predicted land use using model MNL-Cm (right panel)



probabilities as parameters.

In order to evaluate the prediction quality at the grid level, the standard tools cannot be used since there is no reality known at the aggregated level (but only at the point level). We propose to use a weigthed version of the Brier score and the aggregation levels A_0 to A_7 defined previously.

The weighted Brier score is defined for G_g groups ($g \in I_G$, a partition of $\{1, \ldots, n\}$) by:

$$\mathrm{B}_{G_g} = \frac{1}{2n} \sum_{k=1}^{K} \sum_{g \in I_G} \# G_g \Big(\bar{z}_{gk} - \bar{\hat{p}}_{gk} \Big)^2$$

where

$$\bar{z}_{gk} = \frac{1}{\# G_g} \sum_{i \in G_g} z_{ik} \text{ is the observed frequency of land use } k \text{ in group } G_g$$

$$\bar{\hat{p}}_{gk} = \frac{1}{\# G_g} \sum_{i \in G_g} \hat{p}_{ik} \text{ is the estimated probability of land use } k \text{ in group } G_g$$

We recall that the A_0 level denotes the Teruti-Lucas points level, A_1 the segments level, A_2 to A_6 the successive grids and A_7 the global level which is the Midi-Pyrénées region.

Figure 5 (respectively 6) plots the weighted Brier score as a function of the aggregation level for three different prediction or estimated probabilities methods (respectively for five different models). The first method we consider on figure 5 consists in using the estimated probabilities at the points level and aggregate them at the coarser levels. The other two methods have been detailed before and are based on a point prediction which leads to aggregated probabilities at coarser levels. For both figures, the weighted Brier score is quite high at the points level. It decreases substantially already at the segment level and decreases even more at the level A_2 (4 segments). It reaches a stable value at level A_3 (16 segments) and, as expected, takes a zero value for the MNL at the region level.
Figure 5 illustrates the fact that using the estimated probabilities leads to a lower Brier score than the other two aggregation methods. At the points level the multinomial drawing gives particularly bad results but it is already better than the maximum probability prediction at the segment level and is similar to the estimated probabilities method from level A_2.

Figure 5: Weighted Brier scores for model MNL-Cm according to the prediction method

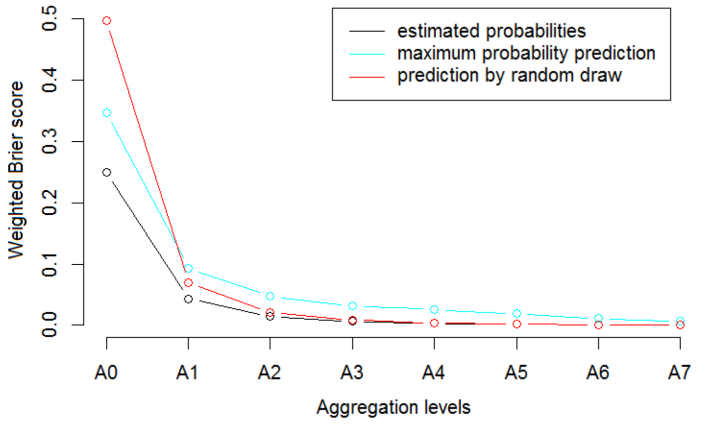


Figure 6: Weighted Brier scores for different models (prediction based on estimated probabilities)

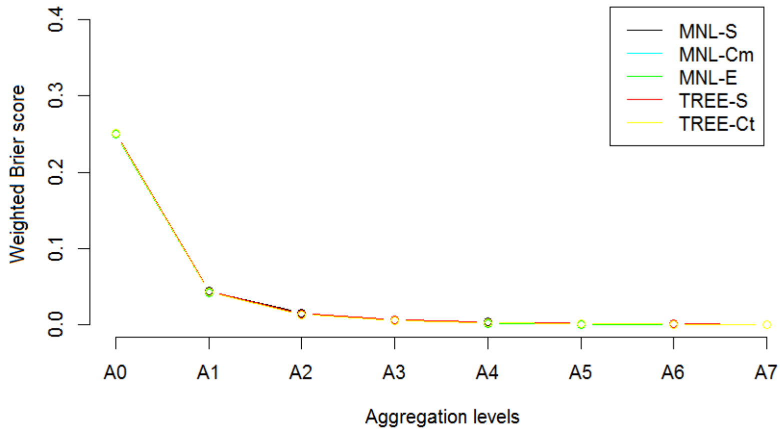


Figure 6 shows that all the previously fitted models lead to equivalent weighted Brier score at all aggregated levels.

Figure 7: Gains obtained by aggregating (total weighted Brier scores, prediction based on estimated probabilities)

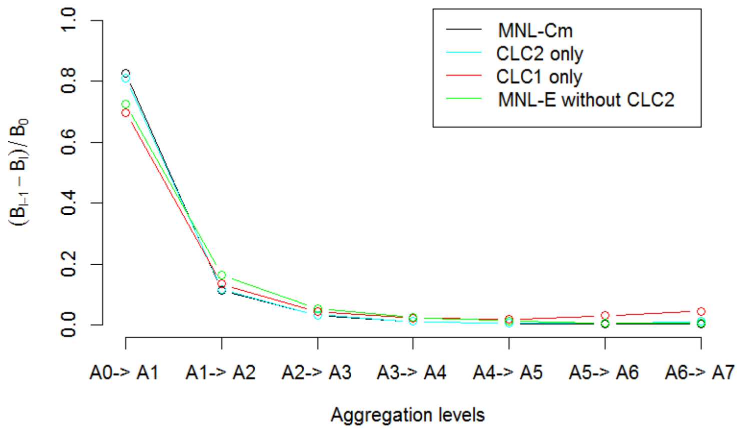


Figure 7 allows to analyze the relative gain in terms of weighted Brier scores between two successive aggregation steps. More precisely we calculate the difference between the score at level A_{l-1} and A_l and divide by the score at the individual level. For model MNL-Cm, this ratio is equal to 0.826 when we go from the individual level to the segment level which means that we gain 83 % on the weighted Brier score by aggregating at the segment level. This plot will help the decision maker for the aggregation level choice. In our application, it shows that it is necessary to aggregate at least at the segment level, which leads to an important gain in terms of prediction quality, while it is not necessary to aggregate at further levels.

6 Conclusion

We aimed at modeling and predicting land use with easily accessible covariates on the Internet or on request from data providers. It is well known that one of the difficulties of modeling land use at the individual level is the frequent lack of "good" explanatory variables or their scale incompatibility. This is particularly true for socio-economic variable (revenue, conversion cost and price) [Chakir, 2015]. We also faced such problems for the socio-economic variables which were only available at the municipalities level or at the NRA (Nouvelles Régions Agricoles) level, but also for the meteorological variables (which resolution is only 25×25 km). Our results demonstrate the high explanatory power of two very simple variables which is the land use information given by Corine Land Cover (at level 2 with 15 categories) and the altitude. These two variables are always selected in our models. Using only these two variables leads already to a correctly classified rate equal to 65.15% ("simple" multinomial logit model). When adding other variables such as meteorological, socio-economic or soil composition variables, the improvement in terms of prediction is only marginal (see Table 4).

We did not take explicitly into account the spatial autocorrelation in our methodology. However in land use modeling spatial autocorrelation can come from two sources [Chakir, 2015]. One possible source is a spatial variable omission that may affect the land use decisions. Another source is the existence of spatial interactions between neighboring parcels (because they belong to the same owner for instance or because they are close to urban areas). Several *ad hoc* procedures can help in reducing the negative potential effects of spatial autocorrelation on the parameters estimation. One possibility is to introduce the latitude and longitude variables as explanatory variables. This approach was not possible in our multinomial logit models because of multicolinearity problems. And these two variables were not selected when using the classification tree methodology. So, following Nelson et al. [2001] and Munroe et al. [2002], we introduced some lagged variables in our models such as the altitude and the socio-economic variables. The lagged altitude is selected in several of our models. So, we may improve on the quality of our models by explicitly taking into account the spatial autocorrelation. However, the estimation of such models is challenging from a computational point of view. The method proposed by Ferdous and Bhat [2013] and Sidharthan and Bhat [2012] seems promising and could be a good alternative to bayesian methods or simulation-based methods that remain quite computationally intensive.

Acknowledgements
This work was supported by the French Agence Nationale de la Recherche through the ModULand project (ANR-11-BSH1-005). We thank the Service de la Statistique et de la Prospective from the Ministère de l'Agriculture in France for giving us access to the data Teruti-Lucas, the Unité INFOSOL from INRA for the data base BDGSF and the Joint Research Center-MARS from the European Commission for the meteorological data (Interpolated Meteorological Data Base). We also thank Antoine Lacroix for the preliminary work done during his first year of Master internship at Toulouse School of Economics.

References

Breiman, L., Friedman, J., Stone, C., and Olshen, R. (1984). *Classification and Regression Trees.* The Wadsworth and Brooks-Cole statistics-probability series. Taylor & Francis.

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly weather review*, 78(1):1–3.

Chakir, R. (2015). L'espace dans les modèles économétriques d'utilisation des sols : enjeux méthodologiques et applications empiriques. *Revue d'Economie Rurale et Urbaine, a paraître.*

Chakir, R., Laurent, T., Ruiz-Gazen, A., Christine, T.-A., and Vignes, C. (2015). Prédiction de l'usage des sols sur un zonage régulier à différentes résolutions et à partir de covariables facilement accessibles. *submitted.*

Ferdous, N. and Bhat, C. (2013). A spatial panel ordered-response model with application to the analysis of urban land-use development intensity patterns. *Journal of Geographical Systems*, 15(1):1–29.

Lubowski, R. N. (2002). *Determinants of land-use transitions in the United States: Econometric analysis of changes among the major land-use categories.* PhD thesis, Harvard University, Cambridge.

McFadden, D. (1974). *Conditional logit analysis of qualitative choice behavior. In: Zarembka P. Frontiers in Econometrics.* Academic Press, New York.

Munroe, D. K., Southworth, J., and Tucker, C. M. (2002). The dynamics of land-cover change in western honduras : Exploring spatial and temporal complexity. *Agricultural Economics*, 27(3):355–369.

Nelson, G. C., Harris, V., and Stone, S. W. (2001). Deforestation, land use, and property rights : Empirical evidence from darien, panama. *Land Economics*, 77(2):187–205.

Sidharthan, R. and Bhat, C. R. (2012). Incorporating spatial dynamics and temporal dependency in land use change models. *Geographical Analysis*, 44(4):321–349. p. 321–349.