Modeling the spatial distribution of crop sequences at a large regional scale using land-cover survey data: A case from France

Ying Xiao, Catherine Mignolet, Jean-François Mari, Marc Benoît

Modeling the spatial distribution of crop sequences at large regional scale using land-cover survey data: a case from France

Ying Xiao[a*], Catherine Mignolet[a], Jean-François Mari[b] and Marc Benoît[a]

[a]INRA SAD UPR 055 ASTER, 662 Avenue Louis Buffet, 88500 Mirecourt, France

[b] Université de Lorraine, LORIA, UMR 7503, 54506 Vandoeuvre-lès-Nancy, France

[*]Correspondingauthor. Address: INRA SAD UPR 055 ASTER, 662 Avenue Louis Buffet, 88500 Mirecourt, France. Tel.: +33 (0)329385500; fax: +33 (0)329385519. E-mail address: yingxiao0201@gmail.com (Y. XIAO).

Abstract:

Assessing the environmental impacts of agricultural production systems requires spatially-explicit information of cropping systems. Projectingchanges in agricultural land use caused by changes in land management practices for analyzing theperformance of land-activities related policies likeagricultural policies alsorequires this type of data as model input. Crop sequences as a vital and widespread adopted agricultural practice are difficult to be directly detected at regional scale.This study presents an innovative stochastic Data Mining aimed at describing the spatial distribution of crop sequences at a large regional scale. The Data Mining is performed by means of Hidden Markov Models and an unsupervised Clustering Analysis that processessequentially-observed(from 1992 to 2003) land-cover survey data of the French mainlandnamedTeruti.The 2549 3-year crop sequences were first identified as major crop sequences across the entire territory including 406 (merged) agricultural districts using Hidden Markov Models. The 406 (merged) agricultural districts were then grouped into twenty-one clusters according to the similarity of the probabilities of occurrences of major 3-year crop sequences by Hierarchical Clustering Analysis. Four cropping systems were further identified: vineyard-based cropping systems, maize monoculture and maize-wheat based cropping systems, temporary pasture and maize-based cropping systems and wheat and barley-based cropping systems. The modeling approach presented in this study provides a tool to extract large-scale cropping patterns from increasingly available time series data of land-cover and use.With this tool, users can (a) identify the homogeneous zone in terms of fixed-length crop sequences across a large territory; (b) understand the characteristics of cropping systems within a region in terms of typical crop sequences; and (c) identify the major crop sequences of a region according to the probabilities of occurrences.

Keywords:

Crop sequences; Cropping patterns; Cropping systems; Hidden Markov models; Agricultural land-use; Teruti survey

1. Introduction

Today, 43% of the area of Europe (Eurostat, 2010) and 36% of the world total area (FAOSTAT, 2011) are dominated by agricultural land use including both cropland and grassland. The current challenge for agronomists, farmers and their allied partners is to satisfy humanity's need for food and fiber as well as the accelerating demand for biomass in an ecologically sustainable way through socially accepted production systems (Miller, 2008).

In the land change science community, over the past decade, the scientific interest of investigating land-cover modification caused by the changes in land management practices has increasingly been noticed by researchers. As pointed out by Lambin et al. (2000), changes in agricultural land use management, e.g., changes in input levels and the effect on profitability, or the periodicity of complex land-use trajectories such as fallow cycles and rotation systems frequently drive land-cover modification. Incorporating into land system models the representation of agricultural land management practices and their changes will improve our understanding of the endogenous driving forces of land-cover modification. Several land system models integrate the module for simulating the farmers' management practice and decision processes (Rounsevell et al., 2003). Agent-based models were specially developed and applied to represent human behavioral and decisional processes in the land system (Matthews et al., 2007). As one of the most significant forms of land-cover modification, agricultural land intensification has recently been studied using different land-use intensity indicators such as livestock density and nitrogen input to UAA (utilized agricultural area) in relation to the land management practices (Herzog et al., 2006). For instance, Temme and Verburg(2011) mapped and modeled agricultural land use intensity in terms of nitrogen input at European Union scale. A multi-scale modeling approach for exploring the spatial-temporal dynamics of European livestock distribution was proposed by Neumann et al. (2011). However, crop rotations as a vital agricultural land management practice are rarely integrated into a land-use modeling framework at regional to global scale (Schönhart et al., 2011).

Crop rotations are defined as the practice of growing a sequence of crops on the same land (Wibberley, 1996). The term 'rotation' implies a cycle and it is characterized by the identified starter crops and the cycle period (e.g. biannual, triennial, 4-years, etc.) (Leteinturier et al., 2006). Because of the multiple benefits of the crop rotations such as increasing crop yields, decreasing the incidence of plant diseases and weeds, maintaining soil fertility, improving the soil structure, preserving biodiversity, crop rotations are a very old widespread practice. In the context of the establishment of new economic, agronomic and governmental policies, farmers will be paid for re-establishing and increasing ecosystem services on agricultural land (Miller, 2008). The positive effect of crop rotations has once more come to the notice of researchers(Merrillet al., 2012; Le Féonet al., 2013).

In the research community which assesses the environmental impacts of agricultural systems, modeling frameworks increasingly incorporated crop rotations instead of single crop for representing cropping patterns. These modeling approaches are related to nitrate leaching in intensive agriculture (Beaudoin et al., 2005), the impacts of agricultural management on the reduction of nitrogen content (Rode et al., 2009), the impact of farming on water resources (Graveline et al., 2012), etc. The manner of representing the cropping systems in terms of crop rotations in these studies was often simplified by expert knowledge based on their own specific field observation or interviews with farmers. A limited number of representative crop rotations were used for describing the cropping patterns in a spatial unit. For allocating these crop rotations within their study area, a crop rotation was usually stochastically assigned to a field, as in the study of Rode et al.

(2009). This simplified approach of representing cropping patterns is due to lack of information about the allocation of crop rotations(Rode et al., 2009). Furthermore, 'crop generator' was proposed for producing spatial and temporal crop distribution under certain conditions such as soil types, agronomic rules or expert knowledge and possiblycalibrated with observed data (Dogliotti et al., 2003; Schönhart et al., 2011). A crop generator was included as an additional module in several hydrological models (Wechsunget al., 2000; Klöckinget al., 2003). The shortcoming of agronomic rules-based crop generators is due to they generate theoretical crop rotations according to the agronomic suitability, but the real crop rotation practices at the field level is influenced by economic condition in the first place, biophysical conditions play only a secondary role (Klöcking et al., 2003). Meanwhile, a study of uncertainty in simulation of nitrate leaching at large regional scale points out the lack of information on the agricultural landuse management presents the greatest uncertainty and underlines its importance (Schmidtet al., 2008). All these reviewed modeling approaches represented cropping patterns from the field to regional meso scale. For representing the cropping patterns at large regional scale or global scale, no modeling work is proposed in the literature. As opposed to the existence of various models at field scale for designing sustainable cropping systems, the lack of cropping system models at regional or global scale results from the unavailability of spatially and temporally explicit information on crop rotations and their associated crop management system (Therond et al., 2011).

The aim of our study is topresent an innovative stochastic Data Mining methodology for describing the spatial distributionof crop sequences at a large regional scale. The Data Mining is performed by means of Hidden Markov Models and an unsupervised Clustering Analysis that processessequentially-observed(from1992 to 2003) land-cover data of the French mainland.

Our study can be considered as an empirical analysis of historical cropping patterns at a large regional scale which will contribute to the scenarios creation of agricultural land-use change caused by changes in land management practices for analyzing the performance of land-activities related policies and land planning.It also provides a tool to extract large-scale spatially-explicit data of cropping patterns from increasingly available time series data of land-cover and use, which will improve the accuracy of the assessment of environmental impacts of agricultural systems. In this study, we define ‘crop sequences’ as the order of appearance of the crops during a fixed period. Crop sequences are strictly synonymous with crop successions. They are the partial or total development of a cycle of rotation or even the basis of several cycles (Leteinturier et al., 2006). As pointed out by the field survey based study, farmers grow different crops over the years in their farm fields without necessarily designing strict rotations (Joannon et al., 2008). For a study of cropping patterns at national scale, we limit our investigation to the major crop sequence related cropping patterns.

We present our modeling approach as follows. First, we describe our study area and the available data source of land-cover. Next, we make a brief introduction of the temporal data mining tool. We then apply our modeling approach, using this historical national land-cover survey data for clustering the French agricultural districts in terms of the similarity of occurrences of crop sequences. Finally, we further characterize the clusters of agricultural districts using both the typically regional crop sequences and the major crop sequences of a region.

2. Materials and Methods

2.1 Study area

Our study area is the French mainland (the island of Corsica is not included) in Western Europe covering 552 thousand square kilometers. Agricultural area as part of the total land area in mainland France was 55.4% in 1992 and 54.2% in 2003 (FAOSTAT, 2011). The area of main agricultural land use at the beginning and end of our study period is described in Table 1. Because of the variation of environmental and socio-economic conditions across the entire territory, the French agricultural production systems reveal their regularity on the spatial distribution. Fig. 1 describes the spatial distribution of farm typology based on the community typology of agricultural holdings in France in 2000 which was carried out by the French Ministry of Agriculture. This EU farm typology is based on economic criteria such as economic size and type of farming. It gives us a glimpse of the spatial distribution of farming systems across French territory. The main cropping zone for cereal and oilseed production is located in central, northern and southwestern France. The livestock zone is situated mainly in the north-west and the Massif Central of France. The mixed cropping and livestock zone is located mainly in southwestern France.

Table 1

Fig. 1

2.2 Data source

The sequential land-cover data used in this study was derived from Teruti databases. Teruti is a two-level sampling survey of land-cover conducted by the French Ministry of Agriculture (Ledoux and Thomas, 1992). Fig. 2 illustrates the sampling method performed in this survey. At the first sampling level, the whole territory was segmented into 4700 grids with an area of 12 × 12 km per grid (Fig. 2a). In most regions, 4 aerial photos among 8 at the positions numbered in 1, 2, 3, 4 (Fig. 2b) were taken within each grid. In total, 15579 aerial photos were taken every June during the survey period. One aerial photo covers around 3.24 square kilometers. At the second sampling level, 36 evenly-spaces sampling points (approximately 300 m apart) were systematically distributedwithin the area of one aerial photo (Fig. 2c). The land-covers of the entire territory were recorded in a matrix in which the sampling points are in a row and the annual records of land-cover in a column. A corpus of 555,382 sampling points labeled with their land-cover during the period from 1992 to 2003 was used in this study.It has detailed information on 81 types of land-cover, including 41 types of crops. Moreover, the Teruti survey provides the constant sampling points which ensure representativeness at different spatial scales based on the occurrences and richness of crops.

Fig. 2

We chose the French agricultural district as the spatial unit in this study. This zoning was established by the French Ministry of Agriculture in 1946 mainly according to the homogeneous agricultural activities and partly the similar environmental conditions such as soil profile and climate (Richard-Schott, 2009). The study by Mignolet et al. (2007) based on interviews with the regional chambers of agriculture indicates that after more than 50 years of development of the socio-technical system, the principal agricultural activities within an agricultural district have remained homogeneous in the Seine Basin. Thus the level of aggregation of Teruti sampling points was defined with respect to the zoning of the agricultural district. All of the 430 agricultural districts in the French mainland territory were incorporated into this study. Because of the small quantity of sampling points (less than 100 points per district) in 21 agricultural districts, we merged them into one of their neighborhood districts according to the similarity of the main land-cover categories. Finally, 406 spatial units including 384 individual agricultural districts and 22 merged agricultural districts were studied.

2.3 Overview of methods

Our strategy of modeling the spatial distribution of crop sequences is to classify the agricultural districts according to the similarity of the occurrences of crop sequences and further to map the result of clustering. The modeling work was carried out in three steps. Firstly, temporal data mining software was applied to estimate the probabilities of the occurrences of crop sequences within each spatial unit. Secondly, we grouped the spatial units in terms of similar crop sequences by performing a classic non-supervised clustering technique. Finally we mapped the result of clustering with the aid of ArcMap 10. In this section we first make a brief introduction of CARROTAGE, our temporal data mining tool used to extract the land-use successions (LUS) in each (merged) agricultural district. We then describe the procedure for identifying the major crop sequences within 406 spatial units using this tool. Finally, the non-supervised classification of the agricultural districts and the cartography of the clustering result will be presented. Here, we take the entire French mainland as a spatial unit for example to demonstrate the procedure of identifying the crop sequences using CARROTAGE. In our analysis, the identification of major crop sequences within a (merged) agricultural district was individually done in the same way for all 406 (merged) agricultural districts.

2.3.1 Description of the temporal data mining tool

CARROTAGE(Le Ber et al., 2006; Mari and Le Ber, 2006), which is a free software,was used to extract the crop sequences on the Teruti survey databases.

Different from several published modeling frameworks of crop sequences which use first-order Markov chains(Aurbacher and Dabbert, 2011; Castellazzi et al., 2008;Salmon-Monviolaet al., 2012), CARROTAGE implements second-order Hidden Markov Models (HMM2). The Hidden Markov Models (HMM) represent the variability inherent to land-cover by means of land-cover distributions organized in a Markov chain rather than representing distinct Markov chains of land-cover. In a HMM2, the Markov chain is a second-order Markov chain that governs the sequence of land-cover distributions. This makes more precise modeling of time events possible, since the land-cover distribution at year *t* depends upon the crop grown in year *t-1* and also *t-2*. Experiment results in speech recognition indicate that HMM2 provides better duration modeling than HMM1 (Mari and Le Ber, 2006). The main feature of HMM of any order is the existence of a learning algorithm (the Baum-Welch algorithm) that can tune the HMM parameters using a corpus of land-cover sequences (the training corpus).

2.3.2 Identification of major land-cover categories within a spatial unit

The first step in data mining is to find an adequate way of encoding the data. We performed a temporal segmentation of the huge matrix of land-cover that covers the period 1992-2003 in order to reduce the number of columns and to represent each sub-period by the distribution of land-cover occurring in this sub-period. Following Le Ber et al. (2006), we specified 12 states left-right HMM2 with one-year land-cover as observation symbol. As our study period covers 12 years, the initial number of states defined for the first specified HMM2 was therefore 12. This HMM2 was trained using the whole matrix and gave 12 land-cover distributions. Among these 12 distributions, many of them were similar. By reducing the number of states, step by step, we got 5 different distributions that defined 5 different land-cover distributions. In this way, crops such as bean, oats, fiber crops, rye, etc. which were not principal crops with extensive growing areas during the whole period but dominant in the territory in several sub-periods, could be incorporated in the study. This procedure of identifying main land-cover using temporal segmentation is useful for us to define which crops will be incorporated into our investigation of crop sequence patterns considering the diversity of crops.

We defined major land-cover types as those types which represented at least 1% of frequency among the total number of land-cover records in the dataset. And all major land-cover types identified in all of the 5 states were then retained as main land-cover categories of a spatial unit for the next analysis of the land-use succession (LUS). Table 2 outlines the main land-cover types identified in these 5 states. Considering the goal of this study was to investigate the crop sequence patterns, we kept crops (except for artificial pasture and temporary pasture) in individual categories and grouped several other land-cover types in one category according to their similarities of characters in land systems (more details see Table 3). Finally, 12 major land-cover categories (Table 3) were defined and were further used for studying LUS.

Table 2

Table 3

2.3.3 Extraction of all LUS involving the major land-cover categories

CARROTAGE allows users to specify HMM2 that can process either single land-cover sequences or sequences made of overlapping fixed length land-cover sub-sequences. For example, the 12 year land-cover sequence: *rapeseed-wheat-barley-rapeseed-wheat-barley*… can be parameterized into a sequence of 11 overlapping 2-year land-cover sub-sequences: *rapeseed-wheat*, *wheat-barley*, *barley-rapeseed*… or even by 10 3-year land-cover sub-sequences: *rapeseed-wheat-barley*, *wheat-barley-rapeseed*, *barley-rapeseed-wheat*… The longer the length of the sub-sequence (say *n*), the more different *n-uplets* we have. This leads to under-training issues when the Baum-Welch algorithm estimates the distributions. On the other hand, the greater *n* is, the more interesting it is for agronomists to find out long crop sequences. In order to choose a suitable observation symbol, we made reference to the previous research work of Le Ber et al. (2006) and Mignolet et al. (2007) in the Seine Basin, where the main field crop cultivation zone in France is located, and to the national statistics published by the French Ministry of Agriculture on farming systems (Agreste, 2010). The former study confirms that crop sequences within the Seine Basin are frequently organized in three or four years. The national agricultural statistics indicate that the crop sequences implemented on French territory generally consist of three times wheat and/or barley and once or twice special regional crops. Considering all the above factors, we choose 3-year land-cover subsequence as the elementary observation symbol in this study.

Referring to the work of Lazrak et al. (2010), we applied a search pattern (Table 4) for extracting all 3-year LUS involving a given major land-cover category. As the field rotation system based on ‘three-field rotation’ and ‘Norfolk four course system’ are widely implemented in Western Europe (Molnar, 2003), we further introduce a field-adopted agronomic rule: starter crop to define the search pattern. The starter crops are often the precedent crop of wheat (mainly) or barley. The field residues of these crops play an important role for soil organic matter and P and K fertilizers restoration. The specialization of starter crops in different agricultural districts constitutes the base of the diversification of cropping patterns while wheat and barley is ubiquitous. Table 4a shows the search pattern we used for extracting the LUS involving these 5 main starter crops in France: maize, rapeseed, peas, sunflower and sugar beet. For the other land-cover categories, the search pattern shown in Table 4b was performed. The introduction of the search patterns in form of ‘starter crop-wheat’ can be considered as a use of HMM2 in a supervised way. In comparison to using one major crop involved search pattern (Lazrak et al., 2010), the search pattern ‘starter crop-wheat’ avoids the repetitions of the same 3-year LUS in different Dirac states (states within HMM2 whose distribution are zero except on a given land-cover category). It keeps the non-agronomical sustainable crop sequences but still implemented in practice like successive cultivation of maize, wheat in a separate state ‘container state’ (state associates to all the other less frequent land-cover categories). It thus gives a better result.

Table 4

One-column ergodic HMM2 (all transitions between states are possible) was performed to carry out this extraction of 3-year land-use successions. The number of Dirac states of model depended on the major land-cover categories previously identified plus a container state (Le Ber et al., 2006).

2.3.4 Filtration of major crop sequences from all 3-year LUS

The goal of this task is to filter out the major 3-year LUS including 3-year successive crops (it means crop sequences in our study) in the output of one-column ergodic HMM2 obtained previously.

We first filtered the 3-year LUS in each Dirac state in the CARROTAGE output files of a spatial unit using double criteria: at least 1% of the probability of occurrence and the appearance of the given land-cover categories in the 3-year LUS. For the container state, all of the LUS which had at least 1% of the probabilities of occurrences were kept for the next step. As the aim of our study was to investigate the major crop sequence related cropping patterns at national scale, a large number of 3-year LUS were removed using the threshold of 1% of the probability of occurrence.

Next, the 406 individual records of main LUS of a (merged) agricultural district were used to build an inventory table in which the 3437 LUS were in a column and the 406 agricultural districts were in a row. In this inventory table, we further removed 888 land-use successions including non-crops in 3-year successions. The remaining 2549 3-year land-use successions, strictly including three successive years of crops, called 'crop sequences' in this study, were retained to cluster 406 (merged) agricultural districts.

Finally, in order to facilitate the interpretation of the characteristics of crop sequence patterns by understanding the context of the agricultural land use, we reclaimed 11 land-use successions which were relevant to the perennial land categories from the 888 removed land-use successions. They were 3-year successions of forest, natural pasture, grass orchard, Alpine meadows, herbaceous vegetation area, rocky areas, water bodies, other semi-natural areas, vegetable gardens and artificial areas with and without construction. Thus, the probabilities of occurrences of 2549 3-year crop sequences and 11 perennial land-covers were retained as the parameter vector of the 406 (merged) agricultural districts.

2.3.5 Clustering and mapping agricultural districts in terms of homogenous crop sequences

In order to cluster the 406 (merged) agricultural districts, we chose the Principal Component methods prior to Ward's Agglomerative Hierarchical Methods (AHC) according to Euclidean distance (Husson et al., 2010) using

R software (R Core Team, 2012) 'FactoMineR' package (Lê et al., 2008). Performing PCA on the raw data is an efficient technique for avoiding high correlations between variables. In our case, taking a typical 3-year 'wheat-barley-rapeseed' crop rotation as an example, the occurrences of its three forms "rapeseed-wheat-barley", "wheat-barley-rapeseed" and "barley-rapeseed-wheat" should be strongly correlated. Thus performing PCA can be considered as a preprocessing of the crop sequence data. It can improve the robustness of the clustering analysis (Josse and Husson, 2012). The PCA was performed without the use of standardization of variables, since the 3-year crop sequences were measured on scales without widely differing ranges and the units of measurement are the same.

In addition, in PCA, 2549 crop sequences were used as active variables and 11 perennial land-covers were used as supplementary variables. The AHCwas performed on the first principal components which account for 80% total inertia. In order to choose the suitable number of clusters in AHC, we first defined the least possible and the most possible number of clusters according to the evident drop in the bar graph of the distance values which was drawn using the package "Cluster" within R. Next, we determinedthe suitable number of clusters within the range of the least and most possible number of clusters with the aid of R software (R Core Team, 2012) 'clValid' package (Brock et al., 2011).All six measures relevant to 'internal' and 'stability' measures implemented in 'clValid' package were used to validate the number of clusters. This number of clusters was then used as argument in the function 'HCPC' of 'FactoMineR' for performing AHC. The advantage of using FactoMineR is that the package integrates a function of the description of clusters by all initial continuous variables both active and supplementary. This measure is named v.test(Lebart et al., 1995),which can be considered as a "standardized" deviation between the mean of those individuals with category *q* and the general average (Husson et al., 2010). In order to understand the characteristics of clusters, the probabilities of occurrences of major 3-year crop sequences were estimated by performing one-column ergodic HMM2 on the corpus of Teruti land-cover data of the agricultural districts belonging to one cluster. The one-column HMM2 contained one Dirac state involving all non-crop land-cover using search pattern (Table 4b).

Finally, the result of clustering analysis was mapped with the aid of ArcMap10 to visualize the crop sequence patterns during 1992-2003.

In addition, while the classification of agricultural districts was established, we further explored the major non-fixed length crop sequences in the territory of one cluster with the aid of the graphic output of one-column ergodicHMM2 (Le Ber et al., 2006).

3. Results

3.1 Descriptive statistical analysis

In PCA, the first two components explained 23.8% and 12.3% of the total inertia, respectively. The first twenty-three principal components which accounted for 80.1% of total variability were used to cluster the agricultural districts. Two-dimensional PCA scores plots and loading plots on PC1 vs. PC2 and PC3 vs. PC4 are shown in Fig. 3. The agricultural districts score plot for PC1 vs. PC2 (Fig. 3a left) reveals two distinguished groups of agricultural districts. One group is projected on the negative dimension of PC1. According to the loading plot of crop sequences (Fig. 3b left), the occurrence of vineyard contributes most to this observed clustering. Another group is projected on the positive dimension of PC2 which correlates with the occurrence of wheat-based crop sequences. In the scores plot of agricultural districts of PC3 vs. PC4 (Fig. 3a right), three groups can be observed. The sugar beet-based crop sequences are heavily loaded for PC4 (Fig. 3b right) which separates the group projected on the negative dimension of PC4 from the others. The second group is projected on the positive dimension of PC4 which can be explained by the sunflower-wheat-based crop sequences having high value of occurrences for PC4 loading.

The occurrence of monoculture of maize is most strongly responsible for the discrimination of one group of agricultural districts that is projected on the positive dimension of PC3. And the occurrence of 3-year fallow partly takes responsibility for this discrimination.

Fig. 3

3.2 Clustering (merged) agricultural districts

At the first step, we used a visual aid, the bar graph of the distance values (Fig. 4) to determine a wide range of the number of clusters. This distance value was the distance value between the two joining clusters that was used by the Ward's method. We looked for the jumps in the decreasing pattern in this bar chart. One possible drop occurs at about the number of clusters = 11 and another occurs at 25. That is, the differences of height between two sizes of clusters after them are all relatively small and about the same size. Next, adopting the cluster validation measures approach implemented in the clValid Package of Brock et al. (2011), we determined the most appropriate number of clusters within the range of 11 to 25. Table 5 shows the result of internal and stability measurements based on different sizes of cluster. Results from the 7 indices indicated that the number of clusters = 21 perhaps 23 was suitable. Considering the tiny differences of the order of crop sequences and their v.test value between the two new small clusters which belonged to the same original cluster, we finally took 21 as the appropriate number of clusters for the AHC. Fig. 5 is the visualization of the result of clusters mapped with ArcMap.

Fig.4

Table 5

Fig. 5

3.3 Description of the crop sequence patterns

The crop sequence patterns delimited in Figure 5 can be described by both the v-test values obtained as outputs of the function HCPC within

FactoMineR and the probabilities of occurrences of major 3-year crop sequences (Table 6).

Table 6

Based on the ten most frequent 3-year crop sequences identified in each cluster, four types of crop sequence patterns can be identified. The first type was vineyard-based cropping systems and it included the clusters 1, 2, 3, 4, 5, 6, 8 and 12. The second type was characterized by the predominance of maize monoculture and maize-wheat-based crop sequences. Clusters 7, 13, 15 and 16 belonged to this type. The third type was temporary pasture and maize-based cropping systems possible for livestock. It included clusters 9, 10 and 11. The fourth type was wheat and barley-based cropping systems including the clusters 14, 17, 18, 19, 20 and 21. This pattern of agricultural districts has been revealed in the previous PCA. Here, we further describe these 21 clusters with the aid of v.test value.

3.3.1 Vineyard-based cropping systems

Four types of vineyard-based cropping systems were distinguishable. The presence of other cropping systems discriminated them. The areas of clusters 1, 2, 4 and 12 were characterized by the predominant mixed systems of vineyard for wine and grape production and other fruit production. Maize monoculture and 3-year successions of sown pastures also occurred in this zone. The differences among these clusters were the occurrences of different fruits which are managed as permanent crop areas. For example peaches and apricots were widely grown in the agricultural districts of cluster 1. Apples, pears and plums were dominant in the zone of cluster 2. Other species of fruits were grown as speciality crops in clusters 4 and 12. Furthermore, monoculture of durum wheat was an important characteristic of the cropping systems of cluster 4. Cluster 3 is the second type of vineyard-based cropping system. Vineyard was absolutely predominant in the agricultural districts of this cluster while maize monoculture and maize-fallow-based crop sequences were also broadly implemented. Clusters 5 and 6 can be identified as the third type of vineyard-based cropping system where vineyards were less frequent than in the zone of cluster 3. And it co-existed with wheat and barley incorporating oilseed crops and sugar beet-based cropping systems. The appearance of beans and artificial pasture based on alfalfa in 3-year crop sequences was a remarkable characteristic of cluster 6. A small cluster (cluster 8) involving 4 agricultural districts was revealed as the fourth type of vineyard-based cropping system. The occurrences of monoculture of durum wheat and other industrial crops discriminated this cluster from the others.

3.3.2 Maize monoculture and maize-wheat-based cropping systems

Maize monoculture was the dominant crop sequence within the agricultural districts of cluster 13. Fallow and vegetables were often integrated into the maize-based crop sequences in this zone. Clusters 7, 15 and 16 belonged to another type of maize-based cropping system. The surface of maize monoculture was important while maize-wheat-based crop sequences and oilseed crops (sunflower and rapeseed)-wheat-based sequences also took a great proportion of growing areas.

3.3.3 Temporary pasture and maize-based cropping systems

Three big clusters 9, 10 and 11 including in total 137 (merged) agricultural districts were characterized by the widespread adoption of successive temporary pasture and temporary maize crop sequences. Maize and wheat-based crop sequences and maize monoculture frequently occurred in the zone of clusters 9 and 10. The high values of v-test of three supplementary variables relevant to the occurrences of rocky areas, alpine meadows and herbaceous vegetation area highlighted that the temporary pasture and maize-based cropping systems in the zone of cluster 11 were probably very extensive and different from the temporary pasture and maize-based cropping systems of clusters 9 and 10. The small cumulative probabilities of occurrences of the 10 most frequent 3-year crop sequences pointed out that arable land under a rotational system occupied a small surface and the extensive area of cluster 11 for agricultural land use was natural permanent grassland.

3.3.4 Wheat and barley-based cropping systems

Six clusters including 115 (merged) agricultural districts belonged to this type of cropping systems. Cluster 14 was the specialist of sunflower cultivation and sunflower was often grown between two years of cereals. The speciality of clusters 17 and 18 was rapeseed. Probably, a typical 3-year “wheat-barley-rapeseed” rotation which consists of three forms: “wheat-barley-rapeseed”, “barley-rapeseed- wheat” and “rapeseed-wheat-barley” was broadly adopted in the zone of these two clusters. We can observe that maize-wheat-based crop sequences occurred frequently in the zone of cluster 17. The presence of 3-year successions of the cultivation of wheat and/or barley discriminated cluster 18 from cluster 17. The “wheat-barley-rapeseed” rotation was also implemented in the zone of cluster 19 and 21. The appearance of pea or sugar beet in 3-year wheat and barley-based crop sequences was an important characteristic of the cropping systems of these two clusters. One remarkable crop sequence that discriminated cluster 21 from 19 is the 3-year sequence of nurseries. The introduction of sugar beet, peas or potatoes between two years of wheat and/or barley was an important characteristic of the cropping systems of cluster 20. The 4-year “wheat-sugar beet- wheat- peas” sequence probably rotated during the study period in the zone of cluster 20.

3.4 Exploration of major non-fixed length crop sequences: example of cluster 17

The major land-cover categories in the thirty agricultural districts of cluster 17 were: wheat, barley, rapeseed, maize, sunflower, temporary pasture, fallow, grassland, other semi-natural zone and perennial areas. One-column ergodicHMM2 with 9 Dirac states and one container state was thus performed. Figure 6 is the graphic output of model in which the probabilities of transitionsbetween two land-cover categories are expressed by the width of the line joining the two land-covers.One can see that, the major crop sequencs are:

(1) Three-year crop rotation “wheat-barley-rapeseed” which consists ofthree 3-year sequences strictly rotating during the whole study period : “barley-rapeseed-wheat” (shown in Fig. 6b by polyline “B1-C2-A3-B4-C5-A6-B7-C8-A9-B10-C11-A12”), “wheat-barley-rapeseed” (polyline “A1-B2-C3-A4-B5-C6-A7-B8-C9-A10-B11-C12”), and “rapeseed-wheat-barley” (polyline “C1-A2-B3-C4-A5-B6-C7-A8-B9-C10-A11-B12”);

(2) Two-year strict crop rotation “maize-wheat” which consists of two rotating 2-year sequences “maize-wheat” (polyline “D1-A2-D3-A4-D5-A6-D7-A8-D9-A10-D11-A12”) and “wheat-maize” (polyline “A1-D2-A3-D4-A5-D6-A7-D8-A9-D10-A11-D12”);

(3) Two-yearcrop rotation “rapeseed-wheat” which consists of two rotating 2-year sequences “rapeseed-wheat” (polyline “C1-A2-C3-A4-C5-A6-C7-A8-C9-A10-C11-A12”) and “wheat-rapeseed” (polyline “A1-C2-A3-C4-A5-C6-A7-C8-A9-C10-A11-C12”);

(4) Monoculture of maize (line D1D12), wheat (line A1A12), and barley (line B5B12);

(5) Long-term fallow (lines F1F4 and F5F10), and temporary pasture (line G1G2);

(6) Two-year sequences “rapeseed-wheat” and “maize-wheat” and one year of wheat may interrupt the predominant 3-year crop rotation “wheat-barley-rapeseed” like “barley-rapeseed-wheat-*rapeseed-wheat*- barley-rapeseed-wheat- barley-rapeseed-wheat-” (polyline “B1-C2-A3-C4-A5-B6-C7-A8-B9-C10-A11-”), “rapeseed-wheat-barley- rapeseed-wheat-barley- rapeseed-wheat-*maize-wheat*-barley-rapeseed” (polyline “C1-A2-B3-C4-A5-B6-C7-A8-D9-A10-B11-C12”) and “wheat-barley-rapeseed-wheat-barley-*wheat*-barley-rapeseed-” (polyline “A1-B2-C3-A4-B5-A6-B7-C8-”), respectively.

One important point has to be noticed is that we can identify the occurrence of major unfixed-length crop sequences, even the exact crop rotations within a spatial unit, but the rate of their occurrences is impossible to be quantified.

Fig. 6

4. Discussion

4.1 A generic approach to describe regional time-space regularities in agricultural landscape

The modeling approach presented in this paper provides a tool to derive spatially-explicit data of cropping patterns at large regional scale from the sequential annual land-cover survey data. With this tool, users can (a) identify the homogeneous zone in terms of fixed-length crop sequences across a large territory, (b) understand the characteristics of cropping systems within a region in terms of typical crop sequences, (c) identify the major crop sequences of a region according to the probabilities of occurrences, and (d) identify the most representative spatial units of each cluster.

The potential application of this modeling approach is as a tool to extract spatially-explicit information on cropping patterns from time series data of land-cover for environmental or economic assessment of agricultural production systems. It can also be used for building historical data of cropping patterns which can be integrated into the land-use change modeling framework for land planning and policy making.

4.2 Limitations of crop sequence- based modeling

The approach proposed here however, has several limitations. These limitations are mainly due to the simplified representation of the complex rotational cropping system. First, we took the concept ‘crop sequences’ which is limited to the order of appearance of the crops during a fixed period instead of the exploration of the exact cycle of crop rotations during the study period. Indeed, most agricultural land management practices are decided at the local scale by the farm holders under different biophysical constraints and socio-economic conditions. Joannon et al. (2008) indicate that farmers grew the crops in a field of their farm over the years without implementing strict crop rotations keeping a degree of freedom in their choices. This may explain why a great number of crop sequences can be observed over a large area. Two observation-based studies confirmed this point. Leteinturier et al. (2006) observed 62499 7-year crop sequences in an area of 255,461 hectares in the Wallonia area of Belgium. In another study in the Central United States, there were 24 crops observed in database and a total of 9,826,083 4-year crop sequences occurred from 2003 to 2010 (Plourde et al., 2013).

Secondly, as we adopted the temporal regularity mining tool based on Hidden Markov Models, we needed to define an observation symbol for the model. In our case, the observation symbol is crop sequence that consists of three components: the length of sequence, the appearance of crops and their order. Our strategy of modeling the spatial distribution of crop sequences is to classify the agricultural districts based on the occurrences of crop sequences within each spatial unit and further mapping the result of clustering. Thus, in order to explore the major crop sequences within each spatial unit, we need to define unique length of sequence for all land units studied. But as we know, in reality, the length of crop rotations ranges from 2 years to 12 years (long crop rotations are often observed in organic farming) (Mudgal and Lavelle, 2010). Hence diversity of crop rotations in terms of the rotation length has been ignored in this study.

Thirdly, based on expert knowledge, we chose 3-year crop sequence as our observation symbol for all 406 (merged) agricultural districts. But the fixed-length crop sequences do not mean a great simplification of complex crop sequences in reality. As monoculture and biennial, triennial and quadrennial crop rotations are widely adopted in the field cropping area for cereal and oilseed production in French mainland. Although this choice of the length of crop sequences may be unable to cover the complete cycles based on the long rotations, biennial, triennial and partly quadrennial rotations covers most areas of arable land. Excepting expert knowledge on local cropping systems, the choice of length of sequence as observation symbol is also limited to both the temporal depth of data available of land-cover and the computing power. Moreover, we kept 2549 major crop sequences for clustering 406 (merged) agricultural districts. Potentially innovative crop sequences with rare occurrences were not specifically taken into account. The more complex cropping patterns involving winter cover crop, intercropping, etc. could not be investigated in this study since the records of the Teruti survey were carried out every June between 1992 and 2003 and each sampling point represents one land-cover type for a year.

4.3 Characteristic of the modeling approach and its potential application to other data source of land-cover

One remarkable characteristic of this modeling approach is the use of historical national land-cover survey data for identifying crop sequences at a large regional scale. One benefit of using this type of survey data of land-cover with detailed information of crops for exploring crop sequences is its time series continuity at the same location. This time series continuity makes it more possible to couple the information of cropping patterns with other statistics on agriculture (i.e., the national census of agriculture, the survey of the structure of agricultural holdings, the survey of agricultural practices) with fewer problems of time mismatch, further improving the description and assessment of the agricultural production systems.

With the development of remote sensing techniques, land-cover data based on the temporal depth of remote sensing imagery is more available. Martínez-Casasnovas et al. (2005) proposed a method of mapping the main multi-year cropping patterns using crop maps which were acquired from supervised classification of Landsat image. The temporal depth of remote sensing imagery is often affected by the quality of the image archive, which suffers reductions of landscape views because of persistent cloud patterns, and changes in the remote sensing system (Rindfuss et al., 2004). Several recent researches make progress in crop classification using time-series remotely sensed data for classifying multiyear agricultural land use or investigating the changes in crop rotation patterns at large regional scale (Wardlowet al., 2007; Brown et al., 2013; Plourdeet al., 2013). Thus, if the remotely sensed multi-temporal land-cover data with maximally detailed land-cover types are available, it is possible to perform our modeling approach to describe the past or current crop sequence patterns from regional to global scale. Ideally, if the data of multi-temporal land-cover covering the entire one year growing season for several years is accessible, it will be possible to explore more complex cropping patterns taking into account both the annual main crops and the cover crops. These high temporal-spatial resolution remote sensing data will provide more spatially-temporally explicit and accurate data for investigating cropping systems.

We emphasize that as the tool we used for extracting crop sequences is a temporal data mining tool, the quality of the corpus of observed sequences strongly influences the model estimation of parameters. Constant and continuous land-cover and use data at the stable location are essential. CARROTAGE is not able to handle the corpus with missing value during the study period, and it is preferable to apply the Hidden Markov Model to large databases.

5. Conclusions

The modeling approach of the spatial distribution of crop sequences presented in this study is an empirical modeling combining a temporal regularity data mining tool based on Hidden Markov Model with a classic unsupervised clustering technique on the annual national land-cover survey dataset. The patterns of crop sequences identified here well represent the homogeneity of the major crop sequences within the zone under similar environmental and socio-economic conditions, as well as the heterogeneity of crop sequence patterns across the entire French mainland territory.

This work allows stakeholders such as advisory services, agencies of agriculture and state agricultural organization to evaluate the state of agricultural land use over a long period. They may therefore evaluate their role, as driving forces, on the state of agricultural production systems.

For future work, two tasks should be carried out: investigating the changes in crop sequence patterns and exploring the determinants of the changes, linking particularly the relationship between farm types (e.g. the economically based EU Community typology for agricultural holdings) and crop sequence patterns.

This modeling approach can be considered as a generic method for modeling the crop sequence patterns using observed land-cover and use data. It is possible to apply it in other cases using other sequential land-cover and use data. It is also possible to perform it at different spatial scales. Regarding the fast growth of investment on the collection of the time series land-cover and use data with categories of crops distinguished by different organizations such as the yearly Land-use/cover area frame statistical survey (LUCAS) funded and launched by Eurostat from 2001, obtaining observed data of cropping patterns becomes possible. However, the large volumes of data of land-cover and use have necessitated the development of innovative data processing and analysis systems for delivering accurate data for global change research.

The contribution of our modeling approach is to extract crop sequence from the sequential land-cover and use dataset to provide spatially-explicit data of cropping patterns for the assessment of environmental impacts of agricultural production systems and modeling the agricultural land-use change under the rotational system.

Acknowledgements

The authors gratefully acknowledge the statistic center SSP of the French Ministry of Agriculture for providing the Teruti land-cover and use dataset and the agricultural holding data of OTEX 2000. We also thank the

Department SAD of INRA and Council of Lorraine for supporting the PhD fellowship of Ying XIAO.The valuable comments of anonymous reviewers are acknowledged.

References

Agreste, 2004. L'utilisation du territoire en 2004 - Nouvelle série 1992 à 2004 Agriculture n° 169, Paris, 83 p.

Agreste, 2010. Grandes cultures-Alternance des cultures, Agreste les Dossiers n° 8, 37 p.URL: 〈http://www.agreste.agriculture.gouv.fr/IMG/file/dossier8_cultures_alternance.pdf〉.

Aurbacher, J., Dabbert, S., 2011. Generating crop sequences in land-use models using maximum entropy and Markov chains. Agric. Syst. 104, 470-479.

Beaudoin, N., Saad, J., Van Laethem, C., Machet, J., Maucorps, J., Mary, B., 2005. Nitrate leaching in intensive agriculture in Northern France: Effect of farming practices, soils and crop rotations. Agric. Ecosyst. Environ. 111, 292-310.

Brock, G., Pihur, V., Datta, S., Datta, S., 2011. clValid: Validation of Clustering Results. R package version 0.6-4.

Brown, J.C., Kastens, J.H., Coutinho, A.C., Victoria, D.d.C., Bishop, C.R., 2013. Classifying multiyear agricultural land use data from Mato Grosso using time-series MODIS vegetation index data. Remote Sens. Environ. 130, 39-50.

Castellazzi, M.S., Wood, G.A., Burgess, P.J., Morris, J., Conrad, K.F., Perry, J.N., 2008. A systematic representation of crop rotations. Agric. Syst. 97, 26-33.

Dogliotti, S., Rossing, W.A.H., van Ittersum, M.K., 2003. ROTAT, a tool for systematically generating crop rotations. Eur. J. Agron. 19, 239-250.

Eurostat, 2010. Results on EU land ocver and use published for the first time, Land Use/Cover Area frame Survey, 4 p.URL:〈http://epp.eurostat.ec.europa.eu/cache/ITY_PUBLIC/5-04102010-BP/EN/5-04102010-BP EN.PDF〉.

FAOSTAT, 2011. Land-use resources domain in FAOSTAT. Food and Agriculture Organization of the United Nations, Rome, Italy. URL: 〈http://faostat.fao.org/site/377/DesktopDefault.aspx?PageID=377#ancor〉 (accessed DATE: 24 May 2012).

Graveline, N., Loubier, S., Gleyses, G., Rinaudo, J.D., 2012. Impact of farming on water resources: Assessing uncertainty with Monte Carlo simulations in a global change context. Agric. Syst. 108, 29-41.

Herzog, F., Steiner, B., Bailey, D., Baudry, J., Billeter, R., Bukácek, R., De Blust, G., De Cock, R., Dirksen, J., Dormann, C.F., De Filippi, R., Frossard, E., Liira, J., Schmidt, T., Stöckli, R., Thenail, C., van Wingerden, W., Bugter, R., 2006. Assessing the intensity of temperate European agriculture at the landscape scale. Eur. J. Agron. 24, 165-181.

Husson, F., Lê, S., Pagès, J., 2010. Exploratory Multivariate Analysis by Example Using R, Computer Science and Data analysis Series. Chapman & Hall/CRC, London.

Joannon, A., Bro, E., Thenail, C., Baudry, J., 2008. Crop patterns and habitat preferences of the grey partridge farmland bird. Agron. Sustain. Dev. 28, 379-387.

Josse, J., Husson, F., 2012. Selecting the number of components in principal component analysis using cross-validation approximations. Comput. Stat. Data Anal. 56, 1869-1879.

Klöcking, B., Strobl, B., Knoblauch, S., Maier, U., Pfutzner, B., Gericke, A., 2003. Development and allocation of land-use scenarios in agriculture for hydrological impact studies. Phys. Chem. Earth 28, 1311-1321.

Lambin, E.F., Rounsevell, M.D.A., Geist, H.J., 2000. Are agricultural land-use models able to predict changes in land-use intensity? Agric. Ecosyst. Environ. 82, 321-331.

Lazrak, E.G., Mari, J.F., Benoît, M., 2010. Landscape regularity modelling for environmental challenges in agriculture. Landsc. Ecol. 25, 169-183.

Le Ber, F., Benoit, M., Schott, C., Mari, J.F., Mignolet, C., 2006. Studying crop sequences with CARROTAGE, a HMM-based data mining software. Ecol. Model. 191, 170-185.

Le Féon, V., Burel, F., Chifflet, R., Henry, M., Ricroch, A., Vaissiere, B.E., Baudry, J., 2013. Solitary bee abundance and species richness in dynamic agricultural landscapes. Agric. Ecosyst. Environ. 166, 94-101.

Lê, S., Josse, J., Husson, F., 2008. FactoMineR: An R package for multivariate analysis. J. Stat. Softw. 25, 1–18.

Lebart, L., Morineau, A., Piron, M., 1995. Statistique exploratoire multidimensionnelle. Dunod, Paris.

Ledoux, M., Thomas, S., 1992. De la photographie aérienne à la production de blé, AGRESTE, la statistique agricole (5).

Leteinturier, B., Herman, J.L., de Longueville, F., Quintin, L., Oger, R., 2006. Adaptation of a crop sequence indicator based on a land parcel management system. Agric. Ecosyst. Environ. 112, 324-334.

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

Martínez-Casasnovas, J.A., Martín-Montero, A., Auxiliadora Casterad, M., 2005. Mapping multi-year cropping patterns in small irrigation districts from time-series analysis of Landsat TM images. Eur. J. Agron. 23, 159-169.

Matthews, R.B., Gilbert, N.G., Roach, A., Polhill, J.G., Gotts, N.M., 2007. Agent-based land-use models: a review of applications. Landsc. Ecol. 22, 1447-1459.

Merrill, S.D., Tanaka, D.L., Liebig, M.A., Krupinsky, J.M., Hanson, J.D., Anderson, R.L., 2012. Sequence effects among crops on alluvial-derived soil compared with those on glacial till-derived soil in the northern Great Plains, USA. Agric. Syst. 107, 1-12.

Mignolet, C., Schott, C., Benoit, M., 2007. Spatial dynamics of farming practices in the Seine basin: Methods for agronomic approaches on a regional scale. Sci. Total Environ. 375, 13-32.

Miller, F.P., 2008. After 10,000 Years of Agriculture, Whither Agronomy? Agron. J. 100, 22-34.

Molnar, I., 2003. Cropping Systems in Eastern Europe: Past, Present, and Future in: Shrestha, A. (Ed.), Cropping Systems: trends and Advances. The Hawworth Press Inc., New York, pp. 623-647.

Mudgal, S., Lavelle, P., 2010. Environmental impacts of different crop rotations in the European Union, Paris, p. 149.URL: 〈http://ec.europa.eu/environment/agriculture/pdf/BIO_crop_rotations%20final%20report_rev%20executive%20summary_.pdf〉

Neumann, K., Verburg, P.H., Elbersen, B., Stehfest, E., Woltjer, G.B., 2011. Multi-scale scenarios of spatial-temporal dynamics in the European livestock sector. Agric. Ecosyst. Environ. 140, 88-101.

Plourde, J.D., Pijanowski, B.C., Pekin, B.K., 2013. Evidence for increased monoculture cropping in the Central United States. Agric. Ecosyst. Environ. 165, 50-59.

R Core Team, 2012. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.ISBN 3-900051-07-0. URL:〈http://www.R-project.org/〉.

Richard-Schott, F., 2009. Le Recensement Général de l'Agriculture de 1955, une référence pour les géographes ? Géocarrefour 84, 271-279.

Rindfuss, R.R., Walsh, S.J., Turner, B., Fox, J., Mishra, V., 2004. Developing a science of land change: Challenges and methodological issues. Proc. Natl. Acad. Sci. U. S. A. 101, 13976.

Rode, M., Thiel, E., Franko, U., Wenk, G., Hesser, F., 2009. Impact of selected agricultural management options on the reduction of nitrogen loads in three representative meso scale catchments in Central Germany. Sci. Total Environ.407, 3459-3472.

Rounsevell, M., Annetts, J., Audsley, E., Mayr, T., Reginster, I., 2003. Modelling the spatial distribution of agricultural land use at the regional scale. Agric. Ecosyst. Environ. 95, 465-479.

Salmon-Monviola, J., Durand, P., Ferchaud, F., Oehler, F., Sorel, L., 2012. Modelling spatial dynamics of cropping systems to assess agricultural practices at the catchment scale. Comput. Electron. Agric. 81, 1-13.

Schönhart, M., Schmid, E., Schneider, U.A., 2011. Crop Rota - A crop rotation model to support integrated land use assessments. Eur. J. Agron. 34, 263-277.

Schmidt, T.G., Franko, U., Meissner, R., 2008. Uncertainties in large-scale analysis of agricultural land use - A case study for simulation of nitrate leaching. Ecol. Model. 217, 174-180.

Temme, A.J.A.M., Verburg, P.H., 2011. Mapping and modelling of changes in agricultural intensity in Europe. Agric. Ecosyst. Environ. 140, 46-56.

Therond, O., Hengsdijk, H., Casellas, E., Wallach, D., Adam, M., Belhouchette, H., Oomen, R., Russell, G., Ewert, F., Bergez, J.-E., Janssen, S., Wery, J., Van Ittersum, M.K., 2011. Using a cropping system model at regional scale: Low-data approaches for crop management information and model calibration. Agric. Ecosyst. Environ. 142, 85-94.

Wardlow, B.D., Egbert, S.L., Kastens, J.H., 2007. Analysis of time-series MODIS 250 m vegetation index data for crop classification in the US Central Great Plains. Remote Sens. Environ. 108, 290-310.

Wechsung, F., Krysanova, V., Flechsig, M., Schaphoff, S., 2000. May land use change reduce the water deficiency problem caused by reduced brown coal mining in the state of Brandenburg? Landsc. Urban Plan. 51, 177-189.

Wibberley, J., 1996. A brief history of rotations, economic considerations and future directions. Aspects of Applied Biology 0265-1491, 1-10.

Table captions:

Table 1

The area of main agricultural landuse in France in 1992 and 2003 (Agreste, 2004).

Table 2

Result of 5 states left-right HMM2: the main land-cover types of French mainland and their percentage of total frequency at five temporal states between 1992 and 2003. This table shows the evolution of several land-covers such as the expansion of forest, the increase in areas of rapeseed cultivation, the decrease in areas of pea cultivation, etc.

Table 3

Major land-cover categories of French mainland and their composition between 1992 and 2003.

Table 4

Search pattern for extracting all 3-year LUS involving one given major land-cover category.

Table 5

Internal and stability measurements on different size of clusters to choose an optimal number of clusters for the dataset.

Table 6

Description of the characteristics of 21 clusters based on the v-test value obtained in AHC and the probabilities of occurrences of the 10 most frequent 3-year crop sequences estimated using one-column ergodic HMM2. Nomenclature used is: A (apples), Ap (apricots), B (barley), Bn (beans), Ch (cherries), Fa (fallow), Fo (nut trees), Fs (berry orchard), G (grassland), H (herbaceous vegetation area), Id (industrial crops), M (maize), N (nursery), O (oats), Oc (other cereals), Ol (oilseed crops), Of (other fodder crops), OS (other semi-natural areas including heathland, moors, hedgerow), Ov (other legumes), P (pea), Pa (artificial pasture sown by alfalfa and clover), Pe (peaches), Pl (plums), Pm (alpine meadows), Pr (pears), Ps (potatoes), Pt(temporary pasture), R (rapeseed), Ry(rye), S (sunflower), Sa (6 major species of fruits and crops), Sb (sugar beet), Ss (mixed orchard of 6 major species), St (rocky areas), Tx (fiber crops), V (vineyards) and W (wheat). CS: crop sequences. AD: agricultural districts. v.test values of variables include both active and supplementary variables.

Figure captions:

Fig. 1.The economic criteria-based EU community typology for agricultural holdings in France in the year 2000. Data supported by the French Ministry of Agriculture.

Fig. 2.Graphical illustration of the two-level sampling method of the Teruti land-cover survey between 1992 and 2003. (a) The entire territory is segmented into 4700 grids. (b) The position of aerial photos taken in each grid. (c) The distribution of 36 sampling points within an aerial photo. One Teruti sampling point covers roughly 100 hectares.

Fig. 3. Principal component analysis based on the occurrences of 3-year crop sequences across 406 (merged) agricultural districts (AD) during 1992-2003. (a) PCA score plots of (merged) agricultural districts. (b) PCA loading plots of 3-year crop sequences. Left: on PC1 vs. PC2. Right: on PC3 vs. PC4. For visibility, only the crop sequences whose squared coefficients of correlation between variable and components > 0.5 for PC1 vs. PC2 and > 0.3 for PC3 vs. PC4 are displayed in plots.

Fig.4. Bar plot of the distance values between the two joining clusters that was used by the Ward's method for hierarchical agglomerative clustering.

Fig. 5. Spatial distribution of 3-year crop sequences in France (overseas departments not included) between 1992 and 2003. Clusters belonging to vineyard-based cropping systems are in the purple series. Clusters belonging to maize monoculture and maize-wheat-based cropping systems are in the orange series. Clusters belonging to temporary pasture and maize-based

cropping systems are in the grey series. Clusters belonging to wheat and barley-based cropping systems are in the green series.

Fig. 6. Graphic output of CARROTAGE. In order to improve the visibility and to guide the audience, we add a grid to give a coordinate for one land-cover in a given year. (a) Original graph: the a posteriori probabilities of transitions between states (diagonal and horizontal lines). Only the transitions whose probability is greater than 0.5% are displayed; (b) Modified graph with adding a grid to give a coordinate for one land-cover in a given year.