# Time-Space Dependencies in Land-Use Successions at the scale of an Agricultural Landscape

***Lazrak, E.G. (1); Benoît, M. (1); Mari, J.F. (2)***

***(1) INRA, UR 055, SAD ASTER - Domaine du Joly, F-88500 Mirecourt, lazrak@mirecourt.inra.fr***

***(2) LORIA, UMR CNRS 7503 et INRIA-Grand Est - B.P. 239, F-54506 Vandœuvre-lès-Nancy***

Abstract: The agricultural landscape can be seen as an assemblage of farm territories. The way farmers organize these territories is a time AND spatial process. Understanding how a land-use succession (LUS) in a parcel depends on LUS of the neighbouring parcels is a milestone to understand the time-spatial organization of the landscape mosaic. In this work, we analyse these time-space dependencies at agricultural landscape scales. We have performed a data mining process based on hidden Markov models (HMM) to identify spatial clusters of similar distributions of LUS in 2 neighbouring parcels, furthermore called cliques. We applied this data mining process to a land-use data set covering the period from 1996 to 2007 of a 350 km² agricultural landscape located within the Niort Plain (France). To take into account the irregular neighbour system of the parcel mosaic, we used a variable depth Hilbert-Peano scan of the area covering the landscape. Through illustrative examples of two contrasted spatial stochastic clusters, we show that considering temporal cliques gives valuable information on the neighbour system in terms of attraction between LUS.

Keywords: HMM2; data mining; temporal cliques

# Introduction

In agricultural landscapes, land-uses are heterogeneously distributed among different agricultural parcels designed by farmers. At a first glance, the landscape spatial organization and its temporal evolution seem both random. Nevertheless, they reveal the presence of logical processes and driving forces related to the soil, climate, cropping system, and economical pressure. The mosaic of parcels together with their soil-occupancies (OCS) can be seen as a noisy picture generated by these different processes. The understanding of how the temporal succession of a parcel influences the neighbouring parcels is a milestone in the data mining process that aims at extracting knowledge from this mosaic. Furthermore, this piece of knowledge is helpful to simulate coherent agricultural landscapes (Le Ber et al., 2009). Recent studies (Le Ber et al., 2006 ; Castellazzi et al., 2008) have shown that the ordered sequences of OCS in each field can be adequately modelled by a Markov process. The OCS at time $t$ depends upon the former OCS at previous times: $t-1$, $t-2$, .... The Markov model or the hidden Markov model (HMM) are able to capture a limited amount of the temporal variability and allow the specification of land-use successions (LUS) in term of which the agricultural landscapes can be described in a more simple way (Lazrak et al., 2009). Similarly, in the spatial domain, the stochastic modelling of situated observations such as OCS or LUS by means of Markov fields is an elegant way to cluster a landscape into homogeneous patches described by probabilistic distributions of the situated observations.

In this work, we process at the same level the temporal and spatial information given by the parcels and their OCS and consider a pair of OCS in 2 neighbouring parcels at time slots $t$ – furthermore called a temporal clique – rather than a single OCS as the basic temporal and spatial information. The stochastic modelling of the temporal cliques allows a spatial and temporal clustering of the landscape and gives valuable informations on the time and spatial dependencies between OCS. Our objective is to develop a generic data mining process, based on HMM and temporal cliques, in order to highlight these time-space dependencies at agricultural landscape scales.

# 1. The land-use database

The case study area is a 350 km² agricultural landscape located within the Niort Plain in Poitou-Charentes region, France. This agricultural landscape has been surveyed for more than 12 years (1996 – 2007). Every year, two land-use surveys (in April and June) allow to monitor both early harvested and late planted crops. These surveys are stored in a GIS geodatabase, in a vector format.

An analysis based on the average frequency of land-uses over the 12-year study period reveals 47 land-uses. These land-uses have been grouped with the help of agricultural experts in 10 categories (table 1) following an approach based on the similarity of crop management.

*Table 1. Composition and average frequencies of adopted land-use categories (Lazrak et al., 2009)*

| Land-use category | Land-use | Cumul. frequency |
|---|---|---|
| Wheat | Wheat, bearded wheat, cereal | 0.337 |
| Sunflower | Sunflower, reygrass followed by sunflower | 0.476 |
| Rapeseed | Rapeseed | 0.600 |
| Urban | Built area, peri-village, road | 0.696 |
| Grassland | Grassland of various types, alfalfa, ... | 0.774 |
| Maize | Maize, ryegrass followed by maize | 0.850 |
| Forest | Forest or hedge, wasteland | 0.884 |
| Winter barley | Winter barley | 0.918 |
| Ryegrass | Ryegrass, ryegrass followed by ryegrass | 0.942 |
| Pea | Pea | 0.964 |
| Others | Spring barley, grape vine, clover, field bean, ryegrass, cereal-legume mixture, garden/market gardening, ... | 1.000 |

# 2. The agricultural landscape mosaic

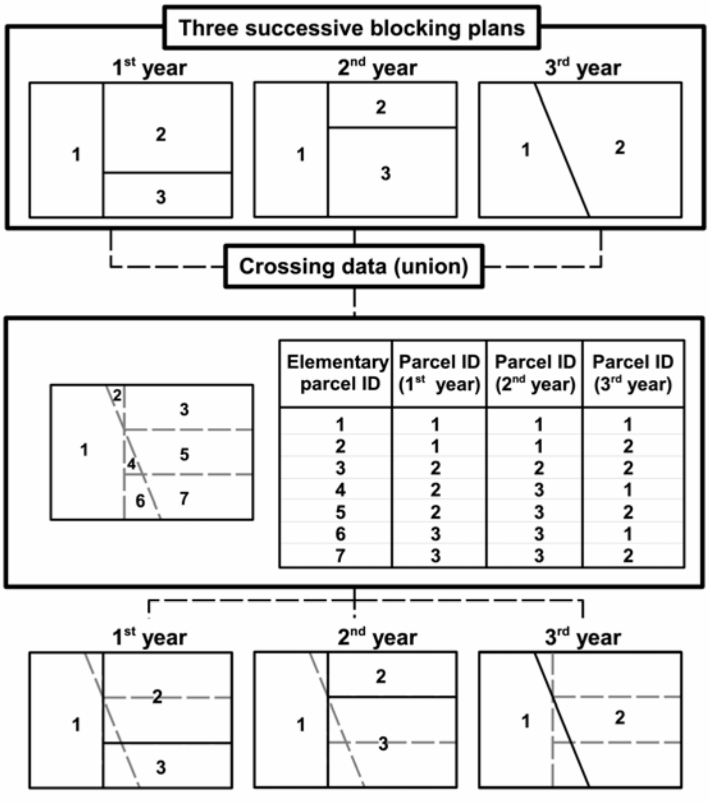
The agricultural landscape can be seen as an assemblage of polygons of variable size – the parcels – where each parcel holds a given OCS.

A parcel can be bounded by a road, a path or a limit of a neighbouring parcel. The parcel boundaries can change every year. To take account of this change, each year, the surveyors update the edges – the boundaries – of parcels in the GIS geodatabase. This led to the definition of the elementary parcel as the result of the spatial union of previous parcel edges (figure 1). There are about 20,000 elementary parcels in the study area over the 1996 – 2007 period. Each elementary parcel holds one succession of OCS during the study period.

The corpus of land-use data is sampled using a regular grid and is represented in a matrix in which the rows represent the land-uses year by year and the lines, the different grid locations.

# 3. Cliques and temporal cliques

Two elementary agricultural parcels represented by 2 polygons are neighbouring if they have at least an edge in common. A clique is a set of parcels in which two unspecified parcels are neighbour. In the mosaic of polygons, the neighbouring relationship – called the neighbour system – is irregular. The parcels have a variable number of neighbours in different geographical directions as opposite to digital images where a site has a fixed number of neighbours. In this paper, we consider simple cliques made of 2 neighbouring parcels represented by the 2 centroids of the parcels. Experimental preliminary results show that the OCS distribution in the cliques is isotropic: the direction defined by the 2 centroids does not carry any information.




| Elementary parcel ID | Parcel ID (1st year) | Parcel ID (2nd year) | Parcel ID (3rd year) |
|---|---|---|---|
| 1 | 1 | 1 | 1 |
| 2 | 1 | 1 | 2 |
| 3 | 2 | 2 | 2 |
| 4 | 2 | 3 | 1 |
| 5 | 2 | 3 | 2 |
| 6 | 3 | 3 | 1 |
| 7 | 3 | 3 | 2 |

*Figure 1. An example of parcel boundary evolution over three successive years. The union of parcel boundaries during this period leads to the definition of seven elementary parcels*

Following Benmiloud and Pieczynski (Benmiloud and Pieczynski, 1995 ; Pieczynski, 2003), we have approximated the Markov field by scanning the 2-D landscape representation with a Hilbert-Peano curve (figure 2). The Markov field is then represented by a Markov chain. To take into account the irregular neighbour system, we have first regularly sampled the area covering the landscape (eg.1 point every 20 m), next have introduced an Hilbert-Peano scanning and finally, have adjusted the fractal depth to the elementary parcel size. The figure 2 illustrates this concept. The sites lying in the same elementary parcel are agglomerated into one point as far they draw the fractal motif. Two successive sites in the $L$-Length fractal curve $(s_{l-1}, s_l)$, $1 \le l < L$, define a clique.

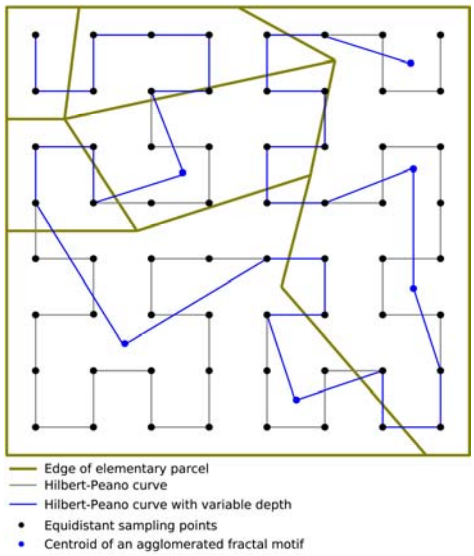



*Figure 2. Variable depth Hilbert-Peano scan to take into account the parcel size. Two successive merging in the bottom left parcel yield to the agglomeration of 16 points*

This scanning introduces a spatial warping and a surface normalization in the parcel mosaic. Large parcels are less sampled, whereas no site agglomeration occurs when the curve crosses the parcel boundaries. The longer is the boundary between two polygons, the more frequent is the clique. Of course, the parcels having singular shape cannot be represented with one centroid and some cliques are situated into the same elementary parcel (figure 2). As a matter of fact, the problem of visiting only once the edges or the vertexes of a graph is known to be NP (non polynomial) hard: there is no algorithm running in a reasonable time to solve it (Rubin, 1974). Our irregular spatial sampling is a crude way to avoid this issue.

The occupancies of a site and its neighbour at time t define the temporal clique. At each site $s_l$ in the variable depth Hilbert-Peano scan, we have defined a feature vector $o_l^t$ with the OCS held in the $(s_{l-1}, s_l)$ cliques:

$$o_l^t = ((s_{l-1}^t, s_l^t), (s_{l-1}^{t+1}, s_l^{t+1})) , \qquad 0 \le t < T-1 , \quad 1 \le l < L \tag{1}$$

where $s_l^t$ is the OCS at time $t$ and index $l$ in the variable depth fractal curve. $t$ is a time index running over the study period, and $l$ the spatial index in the $L$-Length scanning curve. At time $t$, a landscape is then represented by a $(L-1)$-Length sequence of overlapping temporal cliques. We consider also $T-1$ representations to cover the $T$ year length study period due to the overlap artefact.

The cliques inside the same elementary parcel result from the variable depth Hilbert-Peano scan. They are not interesting in the present study. To partially deal with this artefact, feature vectors $o_l^t$ verifying $(s_{l-1}^t = s_l^t)$ and $(s_{l-1}^{t+1} = s_l^{t+1})$ are removed from the resulted distributions.

The feature vector $o_l^t$ is the outcome of 4 random variables $S_t$, $N_t$, $S_{t+1}$ and $N_{t+1}$ that define the observable stochastic process (*cf.* table 2 and table 3).

# 4. The time-space Markovian modelling framework

The way a farmer organizes his territory is a time and spatial process. This time-space dependency becomes more complex at agricultural landscape scales when the agricultural mosaic is built under many farmer's logics. To analyze these dependencies, we rely on 2 assumptions:

1. the OCS of a given field depends upon the OCS of the neighbouring fields (the MRF assumption), and
2. the OCS of a given field in a given year depends also upon the OCS of recent previous years (the Markov chain assumption).

We have modelled the spatial structure of the landscape by a MRF whose sites are random variables of temporal cliques. Like in our previous works (Mari and Le Ber, 2006 ; Lazrak et al., 2009), the MRF has been approximated by a HMM2. This HMM2 has been trained by the EM algorithm on the $T-1$ temporal representations of the landscape.

# 5. The time-space clustering

The stochastic modelling and clustering exhibits patches characterized by distributions of temporal cliques.

- The analysis of rows $S_t$ and $S_{t+1}$ shows the time dependencies at the site level whereas the analysis of rows $N_t$ and $N_{t+1}$ shows the same time dependencies at the neighbour level;
- similarly, the analysis of rows $S_t$ and $N_t$ shows the attraction between OCS;
- furthermore, the joint analysis permits to quantify the attraction between LUS.

Table 2 is a simple example involving the patches tagged as Urban by the stochastic clustering. We can see that the Grassland and Urban categories are stable in the time and have a mutual strong attraction. Less frequent is the neighbourhood occupied by crop successions involving Wheat, Rapeseed and Sunflower.

*Table 2. Temporal cliques in the patches tagged as Urban by the stochastic clustering. Items are listed in decreasing order of frequency.*

| $S_t$ | $N_t$ | $S_{t+1}$ | $N_{t+1}$ |
|---|---|---|---|
| Urban | Grassland | Urban | Grassland |
| Grassland | Urban | Grassland | Urban |
| Sunflower | Urban | Wheat | Urban |
| Urban | Sunflower | Urban | Wheat |
| Urban | Wheat | Urban | Rapeseed |

*Table 3. Temporal cliques in the spatial cluster holding crop successions including Sunflower, Wheat, and Rapeseed. Items are listed in decreasing order of frequency.*

| $S_t$ | $N_t$ | $S_{t+1}$ | $N_{t+1}$ |
|---|---|---|---|
| Wheat | Rapeseed | Rapeseed | Wheat |
| Rapeseed | Wheat | Wheat | Rapeseed |
| Sunflower | Rapeseed | Wheat | Wheat |
| Rapeseed | Sunflower | Wheat | Wheat |
| Wheat | Wheat | Rapeseed | Sunflower |
| Wheat | Wheat | Sunflower | Rapeseed |
| Sunflower | Wheat | Wheat | Rapeseed |
| Wheat | Sunflower | Rapeseed | Wheat |
| Rapeseed | Wheat | Wheat | Sunflower |
| Wheat | Rapeseed | Sunflower | Wheat |

The table 3 is an other example that represents the most frequent items of temporal cliques in the patches holding crop successions including Sunflower, Wheat, and Rapeseed. This table shows clearly that, in theses patches, the OCS located nearby a parcel will be held soon in this parcel. Most likely, this time-space relationship is dictated by the type of crop rotations practiced in this cluster. In fact, a previous data mining study (Lazrak et al., 2009) on the same land-use data base allowed to discover that the main rotations involving Sunflower, Wheat, and Rapeseed in the study area are the quadrennial rotation: (Sunflower-Wheat-Rapeseed-Wheat), and the biennial rotations: (Sunflower-Wheat) and (Rapeseed-Wheat). Furthermore, this spatial cluster describes an open-field agricultural area because the temporal cliques involving either Forest or Grassland in the neighbourhood are not represented.

# Discussion

We have proposed a new representation of agricultural landscapes based on temporal cliques of parcels. To cope with the irregular neighbour system between the parcels, we have specified a variable depth fractal curve that introduces a surface normalization factor and visits the parcels according to their neighbourhood. The sampling becomes irregular and enhances the neighbourhood effects.

Considering temporal cliques rather than single OCS gives a valuable information about the neighbour system between OCS and LUS. This shows the different degree of attraction between LUS in this area and therefore describes the landscape through patches.

Compared to our previous work (Lazrak et al., 2009), the stochastic modelling of the parcel mosaic based on temporal cliques clusters a landscape into agricultural districts that reveal the LUS and the LUS attraction. We put forward the hypothesis that these agricultural districts capture the temporal and spatial variability and can describe, in a simpler way, the agricultural landscapes to achieve a better understanding of the underlying logical processes.

# Acknowledgments

This work was supported by the ANR-ADD-COPT project, the API-ECOGER project and the ANR-BiodivAgrim project. We thank the CNRS team in Chizé for their data records obtained from their "Niort Plain data base".

# References

Benmiloud B., Pieczynski W., 1995. Estimation des paramètres dans les chaînes de Markov cachés et segmentation d'images, *Traitement du signal*, 12(5), p. 433 – 454.

Castellazzi M., Wood G., Burgess P., Morris J., Conrad K.F., Perry J.N., 2008. A systematic representation of crop rotations, *Agricultural Systems*, 97, p. 26–33.

Lazrak E.G., Mari J.-F., Benoît M., 2010. Landscape regularity modelling for environmental challenges in agriculture, *Landscape Ecology*, in press.

Le Ber F., Benoit M., Schott C., Mari J.-F., Mignolet C., 2006. Studying Crop Sequences With CarrotAge, a HMM-Based Data Mining Software, *Ecological Modelling*, 191(1), p. 170-185.

Le Ber F., Lavigne C., Adamczyk K., Angevin F., Colbach N., Mari J.-F., Monod H., 2010. Neutral modelling of agricultural landscapes by tessellation methods – application for gene flow simulation, *Ecological Modelling*, 220(24), p. 3536-3545.

Mari J.-F., Le Ber F., 2006. Temporal and Spatial Data Mining with Second-Order Hidden Markov Models, *Soft Computing*, 10(5), p. 406-414.

Pieczynski W., 2003. Markov models in image processing, *Traitement du signal*, 20(3), p. 255–278.

Rubin F., 1974. A search procedure for hamilton paths and circuits, *Journal of the ACM*, 21(4), p. 576-580.