



HAL
open science

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

► To cite this version:

Ying Xiao, Catherine Mignolet, Jean-François Mari, Marc Benoît. Modeling the spatial distribution of crop sequences at a large regional scale using land-cover survey data: A case from France. *Computers and Electronics in Agriculture*, 2014, 012, pp.51-63. 10.1016/j.compag.2014.01.010 . hal-00967890

HAL Id: hal-00967890

<https://inria.hal.science/hal-00967890>

Submitted on 31 Mar 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

^aINRA SAD UPR 055 ASTER, 662 Avenue Louis Buffet, 88500 Mirecourt, France

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

*Corresponding author. 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).

1 **Abstract:**

2 Assessing the environmental impacts of agricultural production systems
3 requires spatially-explicit information of cropping systems.
4 Projecting changes in agricultural land use caused by changes in land
5 management practices for analyzing the performance of land-activities
6 related policies like agricultural policies also requires this type of data as
7 model input. Crop sequences as a vital and widespread adopted agricultural
8 practice are difficult to be directly detected at regional scale. This study
9 presents an innovative stochastic Data Mining aimed at describing the
10 spatial distribution of crop sequences at a large regional scale. The Data
11 Mining is performed by means of Hidden Markov Models and an
12 unsupervised Clustering Analysis that processes sequentially-observed (from
13 1992 to 2003) land-cover survey data of the French
14 mainland named Teruti. The 2549 3-year crop sequences were first identified
15 as major crop sequences across the entire territory including 406 (merged)
16 agricultural districts using Hidden Markov Models. The 406 (merged)
17 agricultural districts were then grouped into twenty-one clusters according
18 to the similarity of the probabilities of occurrences of major 3-year crop
19 sequences by Hierarchical Clustering Analysis. Four cropping systems were
20 further identified: vineyard-based cropping systems, maize monoculture and
21 maize-wheat based cropping systems, temporary pasture and maize-based
22 cropping systems and wheat and barley-based cropping systems. The
23 modeling approach presented in this study provides a tool to extract large-
24 scale cropping patterns from increasingly available time series data of land-
25 cover and use. With this tool, users can (a) identify the homogeneous zone in
26 terms of fixed-length crop sequences across a large territory; (b) understand
27 the characteristics of cropping systems within a region in terms of typical
28 crop sequences; and (c) identify the major crop sequences of a region
29 according to the probabilities of occurrences.

30 **Keywords:**

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

33 **1. Introduction**

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

41 In the land change science community, over the past decade, the scientific
42 interest of investigating land-cover modification caused by the changes in
43 land management practices has increasingly been noticed by researchers. As
44 pointed out by Lambin et al. (2000), changes in agricultural land use
45 management, e.g., changes in input levels and the effect on profitability, or
46 the periodicity of complex land-use trajectories such as fallow cycles and
47 rotation systems frequently drive land-cover modification. Incorporating
48 into land system models the representation of agricultural land management
49 practices and their changes will improve our understanding of the
50 endogenous driving forces of land-cover modification. Several land system
51 models integrate the module for simulating the farmers' management
52 practice and decision processes (Rounsevell et al., 2003). Agent-based
53 models were specially developed and applied to represent human behavioral
54 and decisional processes in the land system (Matthews et al., 2007). As one
55 of the most significant forms of land-cover modification, agricultural land
56 intensification has recently been studied using different land-use intensity
57 indicators such as livestock density and nitrogen input to UAA (utilized
58 agricultural area) in relation to the land management practices (Herzog et al.,
59 2006). For instance, Temme and Verburg(2011) mapped and modeled
60 agricultural land use intensity in terms of nitrogen input at European Union

61 scale. A multi-scale modeling approach for exploring the spatial-temporal
62 dynamics of European livestock distribution was proposed by Neumann et
63 al. (2011). However, crop rotations as a vital agricultural land management
64 practice are rarely integrated into a land-use modeling framework at
65 regional to global scale (Schönhart et al., 2011).

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

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

91 (2009). This simplified approach of representing cropping patterns is due to
92 lack of information about the allocation of crop rotations(Rode et al., 2009).
93 Furthermore, ‘crop generator’ was proposed for producing spatial and
94 temporal crop distribution under certain conditions such as soil types,
95 agronomic rules or expert knowledge and possiblycalibrated with observed
96 data (Dogliotti et al., 2003; Schönhart et al., 2011). A crop generator was
97 included as an additional module in several hydrological models
98 (Wechsunget al., 2000; Klöcking et al., 2003). The shortcoming of
99 agronomic rules-based crop generators is due to they generate theoretical
100 crop rotations according to the agronomic suitability, but the real crop
101 rotation practices at the field level is influenced by economic condition in
102 the first place, biophysical conditions play only a secondary role (Klöcking
103 et al., 2003). Meanwhile, a study of uncertainty in simulation of nitrate
104 leaching at large regional scale points out the lack of information on the
105 agricultural landuse management presents the greatest uncertainty and
106 underlines its importance (Schmidt et al., 2008). All these reviewed
107 modeling approaches represented cropping patterns from the field to
108 regional meso scale. For representing the cropping patterns at large regional
109 scale or global scale, no modeling work is proposed in the literature. As
110 opposed to the existence of various models at field scale for designing
111 sustainable cropping systems, the lack of cropping system models at
112 regional or global scale results from the unavailability of spatially and
113 temporally explicit information on crop rotations and their associated crop
114 management system (Therond et al., 2011).

115 The aim of our study is to present an innovative stochastic Data Mining
116 methodology for describing the spatial distribution of crop sequences at a
117 large regional scale. The Data Mining is performed by means of Hidden
118 Markov Models and an unsupervised Clustering Analysis that
119 processes sequentially-observed (from 1992 to 2003) land-cover data of the
120 French mainland.

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

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

145 **2. Materials and Methods**

146 **2.1 Study area**

147 Our study area is the French mainland (the island of Corsica is not included)
148 in Western Europe covering 552 thousand square kilometers. Agricultural
149 area as part of the total land area in mainland France was 55.4% in 1992 and
150 54.2% in 2003 (FAOSTAT, 2011). The area of main agricultural land use at

151 the beginning and end of our study period is described in Table 1. Because
152 of the variation of environmental and socio-economic conditions across the
153 entire territory, the French agricultural production systems reveal their
154 regularity on the spatial distribution. Fig. 1 describes the spatial distribution
155 of farm typology based on the community typology of agricultural holdings
156 in France in 2000 which was carried out by the French Ministry of
157 Agriculture. This EU farm typology is based on economic criteria such as
158 economic size and type of farming. It gives us a glimpse of the spatial
159 distribution of farming systems across French territory. The main cropping
160 zone for cereal and oilseed production is located in central, northern and
161 southwestern France. The livestock zone is situated mainly in the north-west
162 and the Massif Central of France. The mixed cropping and livestock zone is
163 located mainly in southwestern France.

164 Table 1

165 Fig. 1

166 2.2 Data source

167 The sequential land-cover data used in this study was derived from Teruti
168 databases. Teruti is a two-level sampling survey of land-cover conducted by
169 the French Ministry of Agriculture (Ledoux and Thomas, 1992). Fig. 2
170 illustrates the sampling method performed in this survey. At the first
171 sampling level, the whole territory was segmented into 4700 grids with an
172 area of 12×12 km per grid (Fig. 2a). In most regions, 4 aerial photos
173 among 8 at the positions numbered in 1, 2, 3, 4 (Fig. 2b) were taken within
174 each grid. In total, 15579 aerial photos were taken every June during the
175 survey period. One aerial photo covers around 3.24 square kilometers. At
176 the second sampling level, 36 evenly-spaces sampling points (approximately
177 300 m apart) were systematically distributed within the area of one aerial
178 photo (Fig. 2c). The land-covers of the entire territory were recorded in a
179 matrix in which the sampling points are in a row and the annual records of
180 land-cover in a column. A corpus of 555,382 sampling points labeled with

181 their land-cover during the period from 1992 to 2003 was used in this
182 study. It has detailed information on 81 types of land-cover, including 41
183 types of crops. Moreover, the Teruti survey provides the constant sampling
184 points which ensure representativeness at different spatial scales based on
185 the occurrences and richness of crops.

186  Fig. 2

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

203 2.3 Overview of methods

204 Our strategy of modeling the spatial distribution of crop sequences is to
205 classify the agricultural districts according to the similarity of the
206 occurrences of crop sequences and further to map the result of clustering.
207 The modeling work was carried out in three steps. Firstly, temporal data
208 mining software was applied to estimate the probabilities of the occurrences
209 of crop sequences within each spatial unit. Secondly, we grouped the spatial
210 units in terms of similar crop sequences by performing a classic non-

211 supervised clustering technique. Finally we mapped the result of clustering
212 with the aid of ArcMap 10. In this section we first make a brief introduction
213 of CARROTAGE, our temporal data mining tool used to extract the land-
214 use successions (LUS) in each (merged) agricultural district. We then
215 describe the procedure for identifying the major crop sequences within 406
216 spatial units using this tool. Finally, the non-supervised classification of the
217 agricultural districts and the cartography of the clustering result will be
218 presented. Here, we take the entire French mainland as a spatial unit for
219 example to demonstrate the procedure of identifying the crop sequences
220 using CARROTAGE. In our analysis, the identification of major crop
221 sequences within a (merged) agricultural district was individually done in
222 the same way for all 406 (merged) agricultural districts.

223 2.3.1 Description of the temporal data mining tool

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

227 Different from several published modeling frameworks of crop sequences
228 which use first-order Markov chains(Aurbacher and Dabbert, 2011;
229 Castellazzi et al., 2008;Salmon-Monviolaet al., 2012), CARROTAGE
230 implements second-order Hidden Markov Models (HMM2). The Hidden
231 Markov Models (HMM) represent the variability inherent to land-cover by
232 means of land-cover distributions organized in a Markov chain rather than
233 representing distinct Markov chains of land-cover. In a HMM2, the Markov
234 chain is a second-order Markov chain that governs the sequence of land-
235 cover distributions. This makes more precise modeling of time events
236 possible, since the land-cover distribution at year t depends upon the crop
237 grown in year $t-1$ and also $t-2$. Experiment results in speech recognition
238 indicate that HMM2 provides better duration modeling than HMM1 (Mari
239 and Le Ber, 2006). The main feature of HMM of any order is the existence

240 of a learning algorithm (the Baum-Welch algorithm) that can tune the HMM
241 parameters using a corpus of land-cover sequences (the training corpus).

242 2.3.2 Identification of major land-cover categories within a spatial unit

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

261 We defined major land-cover types as those types which represented at least
262 1% of frequency among the total number of land-cover records in the
263 dataset. And all major land-cover types identified in all of the 5 states were
264 then retained as main land-cover categories of a spatial unit for the next
265 analysis of the land-use succession (LUS). Table 2 outlines the main land-
266 cover types identified in these 5 states. Considering the goal of this study
267 was to investigate the crop sequence patterns, we kept crops (except for
268 artificial pasture and temporary pasture) in individual categories and
269 grouped several other land-cover types in one category according to their

270 similarities of characters in land systems (more details see Table 3). Finally,
271 12 major land-cover categories (Table 3) were defined and were further
272 used for studying LUS.

273 Table 2

274 Table 3

275 2.3.3 Extraction of all LUS involving the major land-cover categories

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

298 Referring to the work of Lazrak et al. (2010), we applied a search pattern
299 (Table 4) for extracting all 3-year LUS involving a given major land-cover

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

321 Table 4

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

326 2.3.4 Filtration of major crop sequences from all 3-year LUS

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

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

338 Next, the 406 individual records of main LUS of a (merged) agricultural
339 district were used to build an inventory table in which the 3437 LUS were in
340 a column and the 406 agricultural districts were in a row. In this inventory
341 table, we further removed 888 land-use successions including non-crops in
342 3-year successions. The remaining 2549 3-year land-use successions, strictly
343 including three successive years of crops, called 'crop sequences' in this
344 study, were retained to cluster 406 (merged) agricultural districts.

345 Finally, in order to facilitate the interpretation of the characteristics of crop
346 sequence patterns by understanding the context of the agricultural land use,
347 we reclaimed 11 land-use successions which were relevant to the perennial
348 land categories from the 888 removed land-use successions. They were 3-
349 year successions of forest, natural pasture, grass orchard, Alpine meadows,
350 herbaceous vegetation area, rocky areas, water bodies, other semi-natural
351 areas, vegetable gardens and artificial areas with and without construction.

352 Thus, the probabilities of occurrences of 2549 3-year crop sequences and 11
353 perennial land-covers were retained as the parameter vector of the 406
354 (merged) agricultural districts.

355 2.3.5 Clustering and mapping agricultural districts in terms of homogenous
356 crop sequences

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

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

371 In addition, in PCA, 2549 crop sequences were used as active variables and
372 11 perennial land-covers were used as supplementary variables. The
373 AHC was performed on the first principal components which account for 80%
374 total inertia. In order to choose the suitable number of clusters in AHC, we
375 first defined the least possible and the most possible number of clusters
376 according to the evident drop in the bar graph of the distance values which
377 was drawn using the package “Cluster” within R. Next, we determined the
378 suitable number of clusters within the range of the least and most possible
379 number of clusters with the aid of R software (R Core Team, 2012) ‘clValid’
380 package (Brock et al., 2011). All six measures relevant to ‘internal’ and
381 ‘stability’ measures implemented in ‘clValid’ package were used to validate
382 the number of clusters. This number of clusters was then used as argument
383 in the function ‘HCPC’ of ‘FactoMineR’ for performing AHC. The
384 advantage of using FactoMineR is that the package integrates a function of
385 the description of clusters by all initial continuous variables both active and
386 supplementary. This measure is named $v.test$ (Lebart et al., 1995), which can
387 be considered as a “standardized” deviation between the mean of those
388 individuals with category q and the general average (Husson et al., 2010). In
389 order to understand the characteristics of clusters, the probabilities of

390 occurrences of major 3-year crop sequences were estimated by performing
391 one-column ergodic HMM2 on the corpus of Teruti land-cover data of the
392 agricultural districts belonging to one cluster. The one-column HMM2
393 contained one Dirac state involving all non-crop land-cover using search
394 pattern (Table 4b).

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

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

401 **3. Results**

402 3.1 Descriptive statistical analysis

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

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

424 Fig. 3

425 3.2 Clustering (merged) agricultural districts

426 At the first step, we used a visual aid, the bar graph of the distance values
427 (Fig. 4) to determine a wide range of the number of clusters. This distance
428 value was the distance value between the two joining clusters that was used
429 by the Ward's method. We looked for the jumps in the decreasing pattern in
430 this bar chart. One possible drop occurs at about the number of clusters = 11
431 and another occurs at 25. That is, the differences of height between two
432 sizes of clusters after them are all relatively small and about the same size.

433 Next, adopting the cluster validation measures approach implemented in the
434 clValid Package of Brock et al. (2011), we determined the most appropriate
435 number of clusters within the range of 11 to 25. Table 5 shows the result of
436 internal and stability measurements based on different sizes of cluster.
437 Results from the 7 indices indicated that the number of clusters = 21 perhaps
438 23 was suitable. Considering the tiny differences of the order of crop
439 sequences and their v.test value between the two new small clusters which
440 belonged to the same original cluster, we finally took 21 as the appropriate
441 number of clusters for the AHC. Fig. 5 is the visualization of the result of
442 clusters mapped with ArcMap.

443 Fig.4

444 Table 5

445 Fig. 5

446 3.3 Description of the crop sequence patterns

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

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

451 Table 6

452 Based on the ten most frequent 3-year crop sequences identified in each
453 cluster, four types of crop sequence patterns can be identified. The first type
454 was vineyard-based cropping systems and it included the clusters 1, 2, 3, 4,
455 5, 6, 8 and 12. The second type was characterized by the predominance of
456 maize monoculture and maize-wheat-based crop sequences. Clusters 7, 13,
457 15 and 16 belonged to this type. The third type was temporary pasture and
458 maize-based cropping systems possible for livestock. It included clusters 9,
459 10 and 11. The fourth type was wheat and barley-based cropping systems
460 including the clusters 14, 17, 18, 19, 20 and 21. This pattern of agricultural
461 districts has been revealed in the previous PCA. Here, we further describe
462 these 21 clusters with the aid of v.test value.

463 3.3.1 Vineyard-based cropping systems

464 Four types of vineyard-based cropping systems were distinguishable. The
465 presence of other cropping systems discriminated them. The areas of
466 clusters 1, 2, 4 and 12 were characterized by the predominant mixed systems
467 of vineyard for wine and grape production and other fruit production. Maize
468 monoculture and 3-year successions of sown pastures also occurred in this
469 zone. The differences among these clusters were the occurrences of different
470 fruits which are managed as permanent crop areas. For example peaches and
471 apricots were widely grown in the agricultural districts of cluster 1. Apples,
472 pears and plums were dominant in the zone of cluster 2. Other species of
473 fruits were grown as speciality crops in clusters 4 and 12. Furthermore,
474 monoculture of durum wheat was an important characteristic of the cropping
475 systems of cluster 4. Cluster 3 is the second type of vineyard-based cropping
476 system. Vineyard was absolutely predominant in the agricultural districts of
477 this cluster while maize monoculture and maize-fallow-based crop
478 sequences were also broadly implemented. Clusters 5 and 6 can be

479 identified as the third type of vineyard-based cropping system where
480 vineyards were less frequent than in the zone of cluster 3. And it co-existed
481 with wheat and barley incorporating oilseed crops and sugar beet-based
482 cropping systems. The appearance of beans and artificial pasture based on
483 alfalfa in 3-year crop sequences was a remarkable characteristic of cluster 6.
484 A small cluster (cluster 8) involving 4 agricultural districts was revealed as
485 the fourth type of vineyard-based cropping system. The occurrences of
486 monoculture of durum wheat and other industrial crops discriminated this
487 cluster from the others.

488 3.3.2 Maize monoculture and maize-wheat-based cropping systems

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

496 3.3.3 Temporary pasture and maize-based cropping systems

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

509 extensive area of cluster 11 for agricultural land use was natural permanent
510 grassland.

511 3.3.4 Wheat and barley-based cropping systems

512 Six clusters including 115 (merged) agricultural districts belonged to this
513 type of cropping systems. Cluster 14 was the specialist of sunflower
514 cultivation and sunflower was often grown between two years of cereals.

515 The speciality of clusters 17 and 18 was rapeseed. Probably, a typical 3-year
516 “wheat-barley-rapeseed” rotation which consists of three forms: “wheat-
517 barley-rapeseed”, “barley-rapeseed- wheat” and “rapeseed-wheat-barley”
518 was broadly adopted in the zone of these two clusters. We can observe that
519 maize-wheat-based crop sequences occurred frequently in the zone of
520 cluster 17. The presence of 3-year successions of the cultivation of wheat
521 and/or barley discriminated cluster 18 from cluster 17. The “wheat-barley-
522 rapeseed” rotation was also implemented in the zone of cluster 19 and 21.

523 The appearance of pea or sugar beet in 3-year wheat and barley-based crop
524 sequences was an important characteristic of the cropping systems of these
525 two clusters. One remarkable crop sequence that discriminated cluster 21
526 from 19 is the 3-year sequence of nurseries. The introduction of sugar beet,
527 peas or potatoes between two years of wheat and/or barley was an important
528 characteristic of the cropping systems of cluster 20. The 4-year “wheat-
529 sugar beet- wheat- peas” sequence probably rotated during the study period
530 in the zone of cluster 20.

531 3.4 Exploration of major non-fixed length crop sequences: example of 532 cluster 17

533 The major land-cover categories in the thirty agricultural districts of cluster
534 17 were: wheat, barley, rapeseed, maize, sunflower, temporary pasture,
535 fallow, grassland, other semi-natural zone and perennial areas. One-column
536 ergodicHMM2 with 9 Dirac states and one container state was thus
537 performed. Figure 6 is the graphic output of model in which the
538 probabilities of transitions between two land-cover categories are expressed

539 by the width of the line joining the two land-covers. One can see that, the
540 major crop sequences are:

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

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

552 (3) Two-year crop rotation “rapeseed-wheat” which consists of two rotating
553 2-year sequences “rapeseed-wheat” (polyline “C1-A2-C3-A4-C5-A6-
554 C7-A8-C9-A10-C11-A12”) and “wheat-rapeseed” (polyline “A1-C2-
555 A3-C4-A5-C6-A7-C8-A9-C10-A11-C12”);

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

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

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

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

572  Fig. 6

573 **4. Discussion**

574 4.1 A generic approach to describe regional time-space regularities in
575 agricultural landscape

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

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

591 4.2 Limitations of crop sequence- based modeling

592 The approach proposed here however, has several limitations. These
593 limitations are mainly due to the simplified representation of the complex
594 rotational cropping system. First, we took the concept ‘crop sequences’
595 which is limited to the order of appearance of the crops during a fixed
596 period instead of the exploration of the exact cycle of crop rotations during
597 the study period. Indeed, most agricultural land management practices are
598 decided at the local scale by the farm holders under different biophysical

599 constraints and socio-economic conditions. Joannon et al. (2008) indicate
600 that farmers grew the crops in a field of their farm over the years without
601 implementing strict crop rotations keeping a degree of freedom in their
602 choices. This may explain why a great number of crop sequences can be
603 observed over a large area. Two observation-based studies confirmed this
604 point. Leteinturier et al. (2006) observed 62499 7-year crop sequences in an
605 area of 255,461 hectares in the Wallonia area of Belgium. In another study
606 in the Central United States, there were 24 crops observed in database and a
607 total of 9,826,083 4-year crop sequences occurred from 2003 to 2010
608 (Plourde et al., 2013).

609 Secondly, as we adopted the temporal regularity mining tool based on
610 Hidden Markov Models, we needed to define an observation symbol for the
611 model. In our case, the observation symbol is crop sequence that consists of
612 three components: the length of sequence, the appearance of crops and their
613 order. Our strategy of modeling the spatial distribution of crop sequences is
614 to classify the agricultural districts based on the occurrences of crop
615 sequences within each spatial unit and further mapping the result of
616 clustering. Thus, in order to explore the major crop sequences within each
617 spatial unit, we need to define unique length of sequence for all land units
618 studied. But as we know, in reality, the length of crop rotations ranges from
619 2 years to 12 years (long crop rotations are often observed in organic
620 farming) (Mudgal and Lavelle, 2010). Hence diversity of crop rotations in
621 terms of the rotation length has been ignored in this study.

622 Thirdly, based on expert knowledge, we chose 3-year crop sequence as our
623 observation symbol for all 406 (merged) agricultural districts. But the fixed-
624 length crop sequences do not mean a great simplification of complex crop
625 sequences in reality. As monoculture and biennial, triennial and quadrennial
626 crop rotations are widely adopted in the field cropping area for cereal and
627 oilseed production in French mainland. Although this choice of the length of
628 crop sequences may be unable to cover the complete cycles based on the

629 long rotations, biennial, triennial and partly quadrennial rotations covers
630 most areas of arable land. Excepting expert knowledge on local cropping
631 systems, the choice of length of sequence as observation symbol is also
632 limited to both the temporal depth of data available of land-cover and the
633 computing power. Moreover, we kept 2549 major crop sequences for
634 clustering 406 (merged) agricultural districts. Potentially innovative crop
635 sequences with rare occurrences were not specifically taken into account.
636 The more complex cropping patterns involving winter cover crop,
637 intercropping, etc. could not be investigated in this study since the records
638 of the Teruti survey were carried out every June between 1992 and 2003 and
639 each sampling point represents one land-cover type for a year.

640 4.3 Characteristic of the modeling approach and its potential application to
641 other data source of land-cover

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

652 With the development of remote sensing techniques, land-cover data based
653 on the temporal depth of remote sensing imagery is more available.
654 Martínez-Casasnovas et al. (2005) proposed a method of mapping the main
655 multi-year cropping patterns using crop maps which were acquired from
656 supervised classification of Landsat image. The temporal depth of remote
657 sensing imagery is often affected by the quality of the image archive, which
658 suffers reductions of landscape views because of persistent cloud patterns,

659 and changes in the remote sensing system (Rindfuss et al., 2004). Several
660 recent researches make progress in crop classification using time-series
661 remotely sensed data for classifying multiyear agricultural land use or
662 investigating the changes in crop rotation patterns at large regional scale
663 (Wardlow et al., 2007; Brown et al., 2013; Plourde et al., 2013). Thus, if the
664 remotely sensed multi-temporal land-cover data with maximally detailed
665 land-cover types are available, it is possible to perform our modeling
666 approach to describe the past or current crop sequence patterns from
667 regional to global scale. Ideally, if the data of multi-temporal land-cover
668 covering the entire one year growing season for several years is accessible,
669 it will be possible to explore more complex cropping patterns taking into
670 account both the annual main crops and the cover crops. These high
671 temporal-spatial resolution remote sensing data will provide more spatially-
672 temporally explicit and accurate data for investigating cropping systems.
673 We emphasize that as the tool we used for extracting crop sequences is a
674 temporal data mining tool, the quality of the corpus of observed sequences
675 strongly influences the model estimation of parameters. Constant and
676 continuous land-cover and use data at the stable location are essential.
677 CARROTAGE is not able to handle the corpus with missing value during
678 the study period, and it is preferable to apply the Hidden Markov Model to
679 large databases.

680 **5. Conclusions**

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

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

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

698 This modeling approach can be considered as a generic method for
699 modeling the crop sequence patterns using observed land-cover and use data.

700 It is possible to apply it in other cases using other sequential land-cover and
701 use data. It is also possible to perform it at different spatial scales.

702 Regarding the fast growth of investment on the collection of the time series
703 land-cover and use data with categories of crops distinguished by different
704 organizations such as the yearly Land-use/cover area frame statistical survey
705 (LUCAS) funded and launched by Eurostat from 2001, obtaining observed
706 data of cropping patterns becomes possible. However, the large volumes of
707 data of land-cover and use have necessitated the development of innovative
708 data processing and analysis systems for delivering accurate data for global
709 change research.

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

715 **Acknowledgements**

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

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

722 **References**

723 Agreste, 2004. L'utilisation du territoire en 2004 - Nouvelle série 1992 à
724 2004 Agriculture n° 169, Paris, 83 p.

725 Agreste, 2010. Grandes cultures-Alternance des cultures, Agreste les
726 Dossiers n° 8, 37 p. URL:
727 <[http://www.agreste.agriculture.gouv.fr/IMG/file/dossier8_cultures_alternan](http://www.agreste.agriculture.gouv.fr/IMG/file/dossier8_cultures_alternance.pdf)
728 [ce.pdf](http://www.agreste.agriculture.gouv.fr/IMG/file/dossier8_cultures_alternance.pdf)>.

729 Aurbacher, J., Dabbert, S., 2011. Generating crop sequences in land-use
730 models using maximum entropy and Markov chains. Agric. Syst. 104, 470-
731 479.

732 Beaudoin, N., Saad, J., Van Laethem, C., Machet, J., Maucorps, J., Mary, B.,
733 2005. Nitrate leaching in intensive agriculture in Northern France: Effect of
734 farming practices, soils and crop rotations. Agric. Ecosyst. Environ. 111,
735 292-310.

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

738 Brown, J.C., Kastens, J.H., Coutinho, A.C., Victoria, D.d.C., Bishop, C.R.,
739 2013. Classifying multiyear agricultural land use data from Mato Grosso
740 using time-series MODIS vegetation index data. Remote Sens. Environ. 130,
741 39-50.

742 Castellazzi, M.S., Wood, G.A., Burgess, P.J., Morris, J., Conrad, K.F., Perry,
743 J.N., 2008. A systematic representation of crop rotations. Agric. Syst. 97,
744 26-33.

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

747 Eurostat, 2010. Results on EU land cover and use published for the first
748 time, Land Use/Cover Area frame Survey, 4

749 p.URL:<http://epp.eurostat.ec.europa.eu/cache/ITY_PUBLIC/5-04102010-
750 [BP/EN/5-04102010-BP EN.PDF](http://epp.eurostat.ec.europa.eu/cache/ITY_PUBLIC/5-04102010-)>.

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

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

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

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

766 Joannon, A., Bro, E., Thenail, C., Baudry, J., 2008. Crop patterns and
767 habitat preferences of the grey partridge farmland bird. *Agron. Sustain. Dev.*
768 28, 379-387.

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

772 Klöcking, B., Strobl, B., Knoblauch, S., Maier, U., Pfutzner, B., Gericke, A.,
773 2003. Development and allocation of land-use scenarios in agriculture for
774 hydrological impact studies. *Phys. Chem. Earth* 28, 1311-1321.

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

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

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

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

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

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

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

792 Leteinturier, B., Herman, J.L., de Longueville, F., Quintin, L., Oger, R.,
793 2006. Adaptation of a crop sequence indicator based on a land parcel
794 management system. *Agric. Ecosyst. Environ.* 112, 324-334.

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

797 Martínez-Casasnovas, J.A., Martín-Montero, A., Auxiliadora Casterad, M.,
798 2005. Mapping multi-year cropping patterns in small irrigation districts
799 from time-series analysis of Landsat TM images. *Eur. J. Agron.* 23, 159-169.

800 Matthews, R.B., Gilbert, N.G., Roach, A., Polhill, J.G., Gotts, N.M., 2007.
801 Agent-based land-use models: a review of applications. *Landsc. Ecol.* 22,
802 1447-1459.

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

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

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

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

815 Mudgal, S., Lavelle, P., 2010. Environmental impacts of different crop
816 rotations in the European Union, Paris, p. 149.URL:
817 <[http://ec.europa.eu/environment/agriculture/pdf/BIO_crop_rotations%20fin](http://ec.europa.eu/environment/agriculture/pdf/BIO_crop_rotations%20final%20report_rev%20executive%20summary_.pdf)
818 [al%20report_rev%20executive%20summary_.pdf](http://ec.europa.eu/environment/agriculture/pdf/BIO_crop_rotations%20final%20report_rev%20executive%20summary_.pdf)>

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

822 Plourde, J.D., Pijanowski, B.C., Pekin, B.K., 2013. Evidence for increased
823 monoculture cropping in the Central United States. *Agric. Ecosyst. Environ.*
824 165, 50-59.

825 R Core Team, 2012. R: A language and environment for statistical
826 computing. R Foundation for Statistical Computing, Vienna, Austria.ISBN
827 3-900051-07-0. URL:<<http://www.R-project.org/>>.

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

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

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

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

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

843 Schönhart, M., Schmid, E., Schneider, U.A., 2011. Crop Rota - A crop
844 rotation model to support integrated land use assessments. *Eur. J. Agron.* 34,
845 263-277.

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

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

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

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

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

862 Wibberley, J., 1996. A brief history of rotations, economic considerations
863 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.