



# Stochastic and spectral analysis of time series for identifying periodic land use

Jean-François Mari, Odile Horn, Marc Benoît

## ► To cite this version:

Jean-François Mari, Odile Horn, Marc Benoît. Stochastic and spectral analysis of time series for identifying periodic land use. 2020. hal-02944574

**HAL Id: hal-02944574**

**<https://hal.inria.fr/hal-02944574>**

Preprint submitted on 21 Sep 2020

**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.

# Stochastic and spectral analysis of time series for identifying periodic land use

Jean François Mari<sup>a,\*</sup>, Odile Horn<sup>b</sup>, Marc Benoît<sup>c</sup>

<sup>a</sup>*Université de Lorraine, Cnrs, Inria, Loria, F-54000, France*

<sup>b</sup>*Université de Lorraine, LCOMS, ISEA, F- 57000, France*

<sup>c</sup>*Inrae, SAD-ASTER, F-88500, France*

---

## Abstract

This application note shows the interest of the time series auto / cross covariance feature to mine periodic land use / land cover (LU/LC) in an agricultural landscape. A Bayesian point of view has been adopted to cope with the variability of the real-world data sets. The agricultural landscape is represented by a mosaic of plots, each of them holding a time series of annual LU/LC surveys. An agricultural district is seen as a sample drawn from a time series random process. For each LU, the autocovariance function of the random process is first calculated and its Fast Fourier Transform (FFT) performed. The peaks of the power spectrum determine the frequencies – or the corresponding periods – of this LU. In addition, the analysis of the crosscovariance between two LU/LC may exhibit the co-occurrences of these LU/LC at different lags and determine a return time.

In the data mining of agricultural landscapes represented by a mosaic of agricultural plots, results on time series coming from annual LU/LC surveys

---

\*Corresponding author

*Email address:* jfmari@loria.fr (Jean François Mari)

are presented. They show that the auto / cross covariance coefficients and their spectral representation give immediate and accurate information to a data mining analyst on complex land use rotations and their trends for understanding the underlying logical process in landscape dynamics.

*Keywords:*

Periodic crop, land use succession, multivariate random process, time series analysis, autocovariance

---

## 1. Introduction

In agricultural landscapes, the time series of land use and land cover (LU/LC) seem randomly distributed among different agricultural fields (plots) managed by farmers. Nevertheless, they reveal the presence of logical processes and driving forces related to the soil, climate, cropping system, and economical pressure whose understanding is a major challenge mainly for landscape agronomists (Benoît et al., 2012). In these time series, the observation of periodic behavior is of interest for achieving a concise data set representation and extracting knowledge about the underlying logical process. We propose in this application a stochastic framework in which stationary time series of LU/LC coming from annual surveys are analyzed to look for periodic LU and determine crop rotations. A Bayesian point of view has been adopted to cope with the variability of the real-world data sets. The data are assumed to come from a probabilistic source that produces synchronous categorical data. Each agricultural district is seen as a sample drawn from a time series random process.

The originality of our work is twofold: on the one hand, following Berberidis

et al. (2002); Elfeky et al. (2005); Geis and Einax (1996), we propose a method to process time series by means of a spectral analysis of autocovariance coefficients, but we extent their work by using crosscovariance coefficients and categorical data sets. On the other hand, several series of items are handled simultaneously considering them as samples from a random source of categorical variables whose stochastic moments are efficiently estimated when the ergodicity property stands.

We extent our preliminary study on the modeling of the spatial distribution of crop sequences in France (Xiao et al., 2014) by showing on two specific agricultural districts how the autocovariance coefficients and their spectral representation can discern the LU rotations and their trends even in a complex succession scheme.

## 2. Stochastic framework

Let's consider a time sequence  $x_1, x_2, \dots, x_T$  of  $T$  LU/LC drawn from a set  $\mathcal{E} = \{e_1, e_2, \dots, e_K\}$  of  $K$  different labels observed at time  $1, 2, \dots, T$  at a particular plot in an agricultural territory. The  $x_t$  are the realizations of a random variable  $X_t(\omega)$  at time  $t$  in a particular plot represented by  $\omega$ .

Suppose we have logged the LU/LC of a mosaic of plots. The data set defines a matrix  $M$  where row  $\omega$  represents the  $T$  LU/LC of plot  $\omega$  observed at year  $1, 2, \dots, T$ . Column  $t$  represents the LU/LC seen in the whole territory at time  $t$ . This matrix represents a sampling set of the time series  $X_1(\omega), X_2(\omega), \dots, X_T(\omega)$  viewed as a random process.

## 40 2.1. Autocovariance Function

41 In discrete stationary random process, the autocovariance function be-  
42 tween labels  $e_i$  and  $e_j$  is defined as:

$$R_{XX}(\tau)_{[i,j]} = \frac{1}{T} \sum_t \text{Prob}(x_t^i, x_{t+\tau}^j) - E^i E^j \quad (1)$$

43 when  $x_t^i$  means that the LU/LC at time  $t$  is  $e_i$  and  $E^i$  is the mean of label  
44  $e_i$  in the data set  $M$  thanks to the ergodicity property.

45 The probability  $\text{Prob}(x_t^i, x_{t+\tau}^j)$  can be estimated by counting in data set  
46  $M$  the rows where the couple  $(e_i, e_j)$  is seen in columns  $t$  and  $t+\tau$  respectively.

47  $R_{XX}(\tau)_{[i,j]}$  provides information on co-occurrences of  $e_i$  and  $e_j$  at different  
48 lags expressed by  $\tau$ . It can be used as a feature to hypothesize symbol  
49 successions from periodic noisy time series. The result of the spectral analysis  
50 by means of a FFT (Priestley, 1981) of  $R_{XX}(\tau)_{[i,j]}$  as a function of  $\tau$  is a  
51 periodogram whose peaks determine the periods of  $e_i$  (see Fig.2(b)).

## 52 3. Analyzing the return time between crops and meadows in the 53 Vittel area

54 The data used in this study were obtained within the framework of the  
55 research project AGREV 3 (Agriculture Environment Vittel) which accom-  
56 panies Agrivair - a subsidiary of *Nestle water* - in its actions of protection  
57 of natural mineral water. To reach this goal , it is necessary to set up and  
58 to assess agricultural practices which provide a long-term protection of the  
59 resources. As a consequence, crop rotations must be attentively monitored  
60 and *Nestle water* encourages farmers to comply with specifications in which

aid is granted to encourage planting soils in temporary meadows and a crop management without pesticide.

The plot mosaic of the watershed of Vittel Contrexéville was annually investigated between 1996 and 2011. Each parcel has a modality among 11 labels. A time-space clustering of this mosaic has been first performed on the basis on LU/LC successions to specify a sub-territory where the LU/LC successions are modeled by a stationary hidden Markov model by ARPEnt-Age (Mari et al., 2013). The cross covariance was next performed and gave new valuable information on long-term successions by discovering interesting co-occurrences of LU/LC at different lags as shown in Fig. 1.

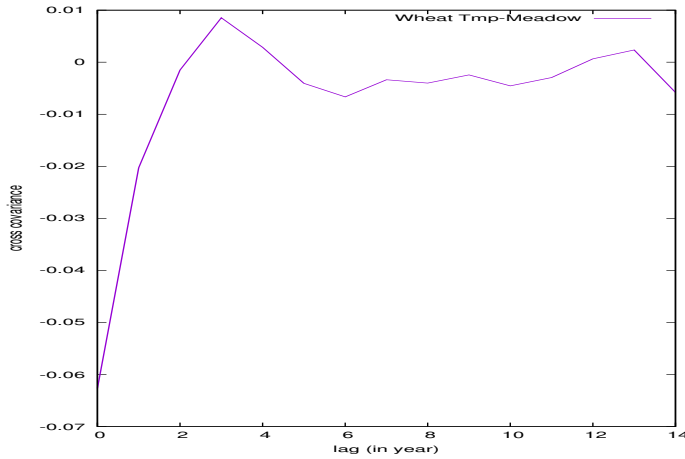


Figure 1: Cross covariance between “Wheat” and “Temporary Meadow” at different lags. The peak located at  $\tau = 3$  is related to the 4-year succession [(“Crop” x 3)-“Temp. Meadow”]. The drop located near  $\tau = 6$  means that a temporary meadow cannot be 6 years later because this LU/LC cannot last more than 5 years

Figure 1 shows an administrative and agronomic feature of the parcels that hold grain crops (Wheat or Barley) in a sub-territory mainly developed by under-contract farmers. In this sub-territory – represented by 42598 16-

length time series – a parcel cannot hold a crop for more than 3 years due to the no-pesticide policy. Therefore a sequence of temporary meadows must be inserted in the mono-culture of crop. This can explain the peak at  $\tau = 3$  because the succession [“Crop” x 3]-”Temp. Meadow”] is frequent. But the “Temp. Meadow” cannot last more than 5 years, otherwise this parcel becomes a permanent meadow and cannot hold another LU. Therefore, before this deadline, these parcels hold another more economically interesting LU, like grain crops. This explains the drop at lag  $\tau = 6$  in the cross covariance function.

#### 4. Analyzing crop sequences in the Saint-Quentinois

For thirty or forty years, increasing human activities (domestic, industrial, agricultural) have gradually degraded the hydro-system of the Seine river, regarding water quality and biological populations.

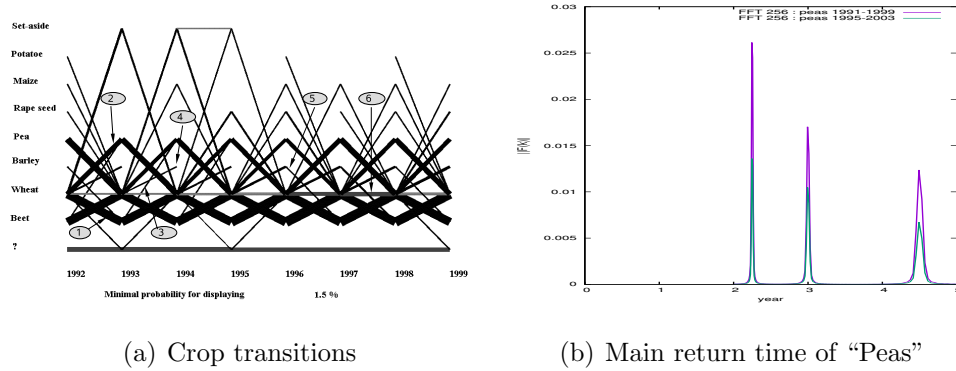


Figure 2: Crop transitions between 1992 and 1999 in the district of Saint-Quentinois (North-east of France) (after Le Ber et al. (2006); Xiao et al. (2014)). The FFT analysis of the autocovariance of “peas” shows 3 periods and a decline of this LU between [1991-1999] and [1995-2003]

87 In this analysis, the autocovariance function and its FFT have been per-  
 88 formed on 2336 time series of annual surveys collected between 1991 and  
 89 2003. Each LU/LC has been classified into eleven modalities. The spectral  
 90 feature can highlight the periods of LU involved in complex rotation scheme.  
 91 The main annual transitions between crops and their trends are displayed in  
 92 figure 2(a) (Le Ber et al., 2006)). The importance of the transition between  
 93 two crops is expressed by the thickness of the line joining the two crops.  
 94 In figure 2(a), the 3-year successions “Wheat-Beet-Wheat” (1) and “Pea-  
 95 Wheat-Pea” (2) seem predominant but no information on longer successions  
 96 is given. A spectral analysis by means of a FFT on the autocovariance of  
 97 “Pea” (Fig. 2(b)) gave a periodogram that immediately showed that this LU  
 98 had 3 periods that can be rounded to: 2, 3, 4 between 1992 and 1999. A  
 99 similar analysis on the time series collected between 1995-2003 showed a  
 100 periodogram with less magnitude indicating a decline of this LU. These in-  
 101 formations have already been given by Le Ber et al. (2006) who performed  
 102 a long and tedious analysis of the triple and quadruple LU in which “Pea”  
 103 was involved. The rotations: “Pea-Wheat”, “Beet-Wheat-Pea” and finally  
 104 “Beet-Wheat-Pea-Wheat” were revealed at that time.

## 105 5. Discussions and conclusions

106 Considering all the LU/LC sequences as a sampling set of a time series  
 107 of random variables  $X_t(\omega)$ , the autocovariance function can be estimated  
 108 efficiently and detects the LU/LC periodicity. In addition, the co-occurrence  
 109 of two LU/LC at different lags can be detected by the cross covariance. In the  
 110 field of agricultural landscapes, this eases the work of an analyst in retrieving

111 periodic sequences or long-term successions for understanding the underlying  
112 logical process in landscape dynamics.

113 In these experiments, we have assumed that the data have two main  
114 properties: stationarity and ergodicity. Stationarity means that the moments  
115 of a time series do not change over the time. In our application, as far as the  
116 driving forces encountered by a farmer on a particular plot do not change,  
117 stationarity of the LU/LC time series can be assumed. Ergodicity is a less  
118 known property that states that the ensemble averages can replace the time  
119 averages. Ergodicity can be assumed in an agricultural landscape where the  
120 crop distribution in the landscape is time independent and corresponds to  
121 the crop distribution in a plot. Then, the time and space delimitation of the  
122 data must be carefully chosen.

## 123 **References**

124 Benoît, M., Rizzo, D., Marraccini, E., Moonen, A.C., Galli, M., Lardon, S.,  
125 Rapey, H., Thenail, C., Bonari, E., 2012. Landscape agronomy: a new  
126 field for addressing agricultural landscape dynamics. *Landscape Ecology*  
127 27, 1385 – 1394.

128 Berberidis, C., Vlahavas, I., Aref, W.G., Atallah, M., Elmagarmid, A.K.,  
129 2002. On the discovery of weak periodicities in large time series, in: Elo-  
130 maa, T., Mannila, H., Toivonen, H. (Eds.), *Principles of Data Mining and*  
131 *Knowledge Discovery*, Springer Berlin Heidelberg, Berlin, Heidelberg. pp.  
132 51–61.

133 Elfeky, M.G., Aref, W.G., Elmagarmid, A.K., 2005. Periodicity detection in

134 time series databases. IEEE Transactions on Knowledge and Data Engi-  
135 neering 17, 875–887.

136 Geis, S., Einax, J., 1996. Multivariate correlation analysis - a method for the  
137 analysis of multidimensional time series in environmental studies. Chemo-  
138 metrics and Intelligent Laboratory Systems 32, 57 – 65.

139 Le Ber, F., Benoît, M., Schott, C., Mari, J.F., Mignolet, C., 2006. Studying  
140 Crop Sequences With CarrotAge, a HMM-Based Data Mining Software.  
141 Ecological Modelling 191, 170 – 185.

142 Mari, J.F., Lazrak, E.G., Benoît, M., 2013. Time space stochastic mod-  
143 elling of agricultural landscapes for environmental issues. Environmental  
144 modelling & software 46, 219–227.

145 Priestley, M., 1981. Spectral Analysis and Time Series. volume 1 and 2.  
146 Academic Press.

147 Xiao, Y., Mignolet, C., Mari, J.F., Benoît, M., 2014. Modeling the spatial  
148 distribution of crop sequences at a large regional scale using land-cover  
149 survey data: a case from France. Computers and Electronics in Agriculture  
150 102, 51–63.

## 151 Appendix A. Spectral analysis of the autocovariance function

152 Let's consider a time sequence  $x_1, x_2, \dots, x_T$  of  $T$  items drawn from a set  
 153  $\mathcal{E} = \{e_1, e_2, \dots, e_K\}$  of  $K$  different labels like the time series made up with  $T$   
 154 land use (LU) observed at time  $1, 2, \dots, T$  at a particular plot in an agricultural  
 155 territory. Each LU/LC belongs to the set  $\mathcal{E}$ . The  $x_t$  are the realizations of  
 156 a random variable  $X_t(\omega)$  at time  $t$  in a particular plot represented by  $\omega$ .  
 157 The time series at a specific plot  $\omega$  is represented by a sequence of  $T$  vectors  
 158 belonging to  $\mathcal{R}^K$ . Vector  $x$  with all components 0 except the  $i$ -th being 1 will  
 159 represent item  $e_i$ . Component  $i$  of vector  $x_t$  is the 1 function – also called  
 160 the Dirac function – that represents the observation of  $e_i$  as a function of  $t$ .  
 161 The time series of items are then parameterized by a series of multivariate  
 162 1 functions allowing processing by the techniques of multivariate random  
 163 signals.

164 When the joint probability distribution  $Prob(x_t, x_{t+\tau})$  is known, the auto-  
 165 / cross second moment is defined by the expectation (Priestley, 1981):

$$\begin{aligned} C_X(\tau) &= E [X_t(\omega) X_{t+\tau}^*(\omega)] \\ &= \frac{1}{T} \sum_t x_t x_{t+\tau}^* Prob(x_t, x_{t+\tau}) \end{aligned} \quad (\text{A.1})$$

166 where  $x^*$  represents the transpose of vector  $x$ .

167 Each term  $x_t x_{t+\tau}^*$  in the sum of Eq. A.1 is a  $K \times K$  matrix. When  
 168  $(x_t, x_{t+\tau})$  represents labels  $(e_i, e_j)$  at time  $(t, t + \tau)$  the product  $x_t x_{t+\tau}^*$  is a  
 169 matrix with 1 at index  $[i, j]$  and 0 everywhere else.

170 Therefore, the general term  $[i, j]$  of  $C_X(\tau)$  is:

$$C_X(\tau)[i, j] = \frac{1}{T} \sum_t Prob(x_t^i, x_{t+\tau}^j) \quad (\text{A.2})$$

171 where  $x_t^i$  is the component  $i$  of vector  $x_t$  and means that item  $e_i$  has been  
 172 observed at time  $t$ .

173 On the basis of Equation A.1, the autocovariance function is defined as  
 174 an auto cross second moment with centered random variables.

$$R_{XX}(\tau) = E[X_t(\omega)X_{t+\tau}^*(\omega)] - E[X_t(\omega)]E[X_{t+\tau}^*(\omega)] \quad (\text{A.3})$$

175 If a stochastic process is stationary, these statistic moments are time  
 176 invariant. In that case:

$$E[X_t(\omega)] = E[X_{t+\tau}(\omega)] = E[X(\omega)]$$

177 The general term  $[i, j]$  is equal to:

$$R_{XX}(\tau)[i, j] = \frac{1}{T} \sum_t \text{Prob}(x_t^i, x_{t+\tau}^j) - E^i E^j \quad (\text{A.4})$$

178 where  $E^i$  is component  $i$  of vector  $E[X(\omega)]$ .

179 The frequencies that compose the autocovariance function  $R_{XX}(\tau)[i, j]$   
 180 as a function of  $\tau$  may be extracted by a discrete Fourier transform (FFT for  
 181 real valued function or DFT in the case of discrete data).

182  $R_{XX}(\tau)[i, j]$  is first windowed with a N-length Hamming window to alle-  
 183 viate the artifacts related to the limited nature of the sequence.

184 Let's denote  $f_{i,j}(\tau) = R_{XX}(\tau)[i, j]$  a function of  $\tau$  whose parameters are  
 185  $i$  and  $j$ , the DFT  $\mathcal{F}_{i,j}(k)$  on  $N$  points is defined by:

$$\mathcal{F}_{i,j}(k) = \sum_{\tau=0}^{N-1} f_{i,j}(\tau) \exp -j \frac{2\pi k \tau}{N} \quad k = 0, \dots, N-1 \quad (\text{A.5})$$

186 the values  $\frac{k}{N}$  are called the frequencies whereas  $\frac{N}{k}$  represents the periods.  
 187 Let's denote  $\|\mathcal{F}_{i,j}(k)\|$  the modulus of the complex number  $\mathcal{F}_{i,j}(k)$ . The  
 188 values  $\|\mathcal{F}_{i,j}(k)\|^2$ ,  $k = 0, \frac{N-1}{2}$  define the power spectrum.

189      The graphical representation of the power spectrum as a function of  $k/N$   
190 is called a spectrogram and viewed as a function of  $N/k$  a periodogram.

191      To get a reasonable frequency resolution, it is possible to use longer win-  
192 dows that are padded with zero. In our work, due to the limited length of  
193 the sequences (roughly between 10 and 20 items), the sequences to fill the  
194 DFT window have been duplicated and padded with zeros.

195      If the hypothesis of a stationary process is unrealistic, it is possible to  
196 define a narrower sliding window that will be shifted along the data set.