



HAL
open science

Local Factorial Analysis of Time Series.

Ahlame Douzal Chouakria, Nacer Hammami, Catherine Garbay

► **To cite this version:**

Ahlame Douzal Chouakria, Nacer Hammami, Catherine Garbay. Local Factorial Analysis of Time Series.. International Statistical Institute, 2007, Portugal. hal-00360489

HAL Id: hal-00360489

<https://hal.science/hal-00360489>

Submitted on 11 Feb 2009

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.

Local Factorial Analysis of Multivariate Time Series

Douzal Chouakria, Ahlame

TIMC-IMAG-CNRS (UMR 5525), Université Joseph Fourier Grenoble 1,

F-38706 LA TRONCHE Cedex, France

E-mail: Ahlame.Douzal@imag.fr

Hammami, Nacer

TIMC-IMAG-CNRS (UMR 5525), Université Joseph Fourier Grenoble 1,

F-38706 LA TRONCHE Cedex, France

E-mail: Nacer.Hammami@imag.fr

Garbay, Catherine

CLIPS-IMAG, Université Joseph Fourier Grenoble 1,

385 Rue de la Bibliothèque BP53 38041 Grenoble Cedex 9

E-mail: Catherine.Garbay@imag.fr

1 Introduction

It's quite often that we are faced to datasets where *a priori* relationship structure is defined on the statistical units. Such structure is generally represented by a neighborhood graph, where vertices are statistical units and the edges connect neighboring units. A pioneering work to include *a priori* structure in factorial analysis was proposed by Lebart (Lebart (1969) and Banet et al. (1984)) with the local and contiguity analysis which mainly consists in decomposing the total variance through the neighboring and non neighboring units. Many other studies follow, for instance, Le Foll (1982) generalizes the local analysis to weighting neighborhood graphs. Mom (1988) proposes a new operator generalizing the discriminant factorial analysis to account for *a priori* neighboring structure. Escofier and Benali (1989) propose an interesting work on factorial smooth analysis and factorial analysis of local differences. Thioulouse et al. (1995) propose another decomposition of the total variance into local variance and global covariance where units are weighted and centered according to the neighboring relationship structure.

This paper focuses on the local analysis for the exploratory of multivariate time series partitioned into *a priori* classes. The studied structure includes two kinds of relationship: the temporal relationship between observations of a same time series, and the membership relationship between time series belonging to a same class. According to that particular structure, we explore two local analysis approaches based on two distinct neighborhood matrices.

2 Local Analysis

2.0.1 Notation

Let $X = [x_{ij}]$ be an $(n \times p)$ matrix describing n individuals by p variables. A neighboring structure is defined on the set of n individuals, it is represented by a neighborhood graph, where vertices identify individuals and edges connect neighboring individuals. Let $M = [m_{ij}]$ be a symmetric $(n \times n)$ matrix describing the neighboring relationship between the n individuals: $m_{ij} = 1$ if i and j are neighbors; 0 otherwise. $m_{i,i}$ is assumed equal to 0. Let $N = \text{diag}(n_1, \dots, n_n)$ be a diagonal matrix where $n_i = \sum_{j=1}^n m_{ij}$ defines the number of i 's neighbors. We consider $m = \sum_{i,j} m_{ij} = \sum_{i=1}^n n_i$ as the total number of couple of neighbors.

2.0.2 Local variance/covariance matrix

Let $V = [v_{jj'}]$ be the classical empirical covariance matrix of X . To include the neighboring structure Lebart proposes a decomposition of the covariance matrix into two covariance matrices: the first one, called local covariance, is defined on the neighboring individuals and the second one is defined on the non neighboring individuals. The local covariance matrix $V_L = [v_{jj'}^L]$ expression is:

$$V_L = \frac{1}{2m} X'(N - M)X \quad \text{with} \quad v_{jj'}^L = \frac{1}{2m} \sum_{i, i' \text{ neighbors}} (x_{ij} - x_{i'j})(x_{ij'} - x_{i'j'})$$

Note that V_L is identical to V in the case of a complete structure where any couple (i, i') are neighbors. Local analysis consists of the diagonalization of V_L to look for the first principal components maximizing the local variance. On the contrary, the diagonalization of $\frac{V}{V}$ to look for the axis minimizing the local variance while maximizing the variance due to the non neighboring individuals defines an extension of the discriminant factorial analysis to include *a priori* structure. It becomes the classical discriminant factorial analysis in the case of a neighborhood graph composed of disjoint cliques where edges connect individuals belonging to a same class.

3 Exploratory of multivariate time series

Let's consider a matrix X describing n multivariate time series S_1, \dots, S_n through p variables X_1, \dots, X_p observed at the instants of time t_1, \dots, t_k . We assume the set of time series partitioned into C *a priori* classes, and $C(S_i)$ denoting the S_i 's class.

3.0.3 Neighboring structure

According to the above dataset X , we distinguish two *a priori* relationships: the temporal relationship between the observations of a same time series, and the membership relationship between the time series belonging to a same class. This structure can be represented by a neighborhood hyper-graph, where an hyper-node, representing a time series, is described by a neighborhood temporal graph connecting the observations of a same time series, and an hyper-edge, connecting two neighboring time series, is described by a set of edges connecting the observations of the two time series in question. Let's define the general form of the corresponding neighborhood block matrix $M = [M_{ij}]$:

$$M = \begin{matrix} & S_1 & \dots & S_n \\ S_1 & \left[\begin{array}{ccc} M_{11} & \dots & M_{1n} \\ \vdots & \dots & \vdots \\ M_{n1} & \dots & M_{nn} \end{array} \right] \\ \vdots & & & \\ S_n & & & \end{matrix}$$

where $M_{ij} = [m_{ls}]$ ($i \neq j$) is the $(k \times k)$ symmetric matrix defining the membership relationship between the time series S_i and S_j , and $M_{ii} = [m_{ls}]$ the symmetric matrix defining the temporal relationship between the observations of S_i .

3.0.4 Exploratory analysis purposes

In this paper we propose two exploratory analysis of the set of multivariate time series. The first exploratory analysis purpose is to look for the first axis best separating the time series belonging to different classes; while making closer the time series of a same class. These axis should give back of the central position of the time series not of their temporal behavior. The second exploratory analysis, similarly to the first analysis, looks for the first axis best separating the time series belonging to different classes, while making closer the time series of the same class, but with preserving the temporal behavior characterizing each time series.

With respect to the first exploratory analysis purpose, we propose the following neighborhood block matrices $M = [M_{ij}]$ definition:

$$M_{ij} = \begin{cases} \text{zero matrix} & \text{if } C(S_i) = C(S_j) \\ \text{unit matrix} & \text{if } C(S_i) \neq C(S_j) \end{cases}$$

Let V_L be the local covariance matrix based on the neighborhood matrix M . The first analysis, consists in diagonalizing $\frac{V_L}{V}$ to look for the first principal components maximizing the local variance. This local analysis will reveal the main directions separating well the time series belonging to different classes, while bringing closer the time series of a same class and the observations of a same time series. As M_{ii} is a zero matrix, factorial projections will reveal some central position of each time series instead of their behaviors. According to the second exploratory analysis purpose, let's consider the following neighborhood block matrices $M_1 = [M_{ij}^1]$ and $M_2 = [M_{ij}^2]$ definitions:

$$M_{ij}^1 = \begin{cases} \text{zero matrix} & \text{if } C(S_i) = C(S_j) \text{ and } (i \neq j) \\ m_{ls} = 1 & \text{if } i = j \text{ and } s = l \pm 1; \quad 0 \text{ otherwise} \\ \text{unit matrix} & \text{if } C(S_i) \neq C(S_j) \end{cases}$$

$$M_{ij}^2 = \begin{cases} \text{Identity matrix} & \text{if } C(S_i) = C(S_j) \text{ and } (i \neq j) \\ \text{zero matrix} & \text{if } i = j \\ \text{zero matrix} & \text{if } C(S_i) \neq C(S_j) \end{cases}$$

Let V_L^1, V_L^2 be the local covariance matrices based on the neighborhood matrices M_1 and M_2 respectively. To bring out as well as possible the temporal behavior of each time series, we propose to maximize the local covariance V_L^1 defined for one on the neighboring temporal observations, and to minimize the local variance V_L^2 defined by the time series of a same class, based on their values observed at the same time. Thus, the second exploratory analysis, consists in diagonalizing $\frac{V_L^1}{V_L^1 + V_L^2}$ to look for the first principal components maximizing the local variance V_L^1 while minimizing the local variance V_L^2 .

4 Application and results

To illustrate the above two variants of the discriminant factorial analysis, we consider the Auslan dataset (<http://kdd.ics.uci.edu/databases/auslan/auslan.data.html>) describing a set of hand signs language of the deaf Auslan community. Instances of the hand signs were collected using an instrumented glove. Position information is calculated on the basis of ultrasound emissions from emitters on the glove. Each sign is measured through a time series described by 8 variables: x (left/right), y (up/down), z (backward/forward), and roll (the palm pointing up or down) giving the spatial position of the glove, and the four remaining variables thumb, fore middle and ring finger indicating if a finger is fully bent or not. For clarity reasons, we have limited our dataset to the 5 signs "Eat", "Boy", "You", "Wild" and "Pen" of the Adam's session 2, with 6 instances per sign. The 30 collected time series were resampled to 57 points (the maximal observed length). We have processed the two above exploratory analysis on the 30 time series. Figure 1 illustrates, on the left the coordinates on the first factorial plan of the 30 time series according to the first exploratory analysis, and on the right the factorial coordinates obtained through the second exploratory analysis. We use a three-dimensional visualization to show the temporal behavior of the projected time series. We can see that both exploratory analysis succeed to separate well the 5 time series classes, with projecting very closely the time series of each class. Let's focus on the main differences between the two produced factorial plans. Based on the temporal neighboring relationship, the local variance of a projected time series is maximum if the behavior of the projected time series is as close as possible to the initial time series behavior (i.e. behavior in the initial descriptive space). On the contrary, the local variance, of a projected time series is minimum if the projection sub-space is some what orthogonal to the initial time series behavior, where all the

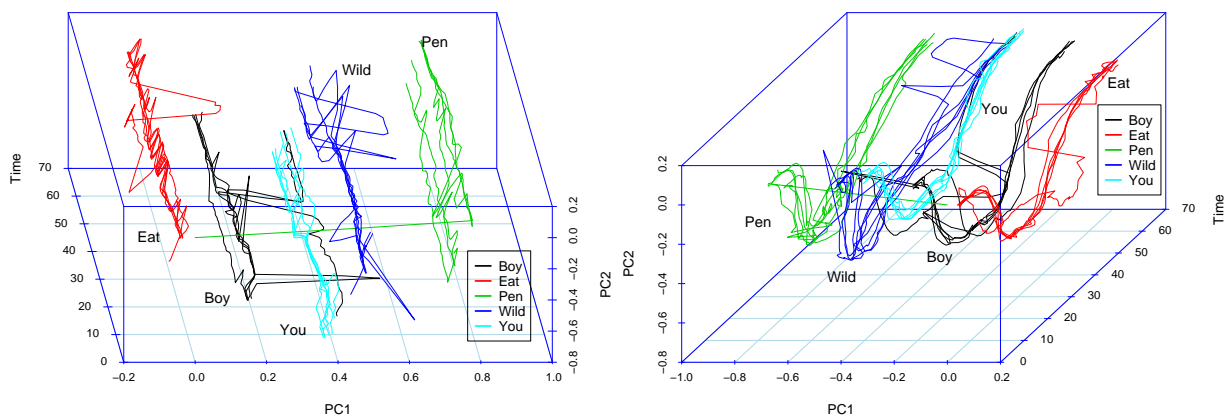


Figure 1: First principal components of the 30 time series

values are projected in a minimal variance domain. Indeed, through the first exploratory analysis, looking for axis minimizing the local temporal variance, the obtained factorial plan (figure 1 on the left) shows well that each projected time series is of minimal variability, all time series are roughly described by a lines, each giving the central position of a projected time series. Whereas, through the second exploratory analysis maximizing the local temporal variance, the obtained factorial plan (figure 1 on the right) shows well the two dimensional behavior characterizing each time series classes.

5 Conclusion

This paper focuses on the exploratory analysis of a set of multivariate time series partitioned into *a priori* classes. We propose two exploratory analysis. The both exploratory analysis look for a sub-space separating well the time series of different classes; while making closer those belonging to a same class. The main difference, is that the first exploratory analysis looks for a subspace working out the main central position of each time series regardless to their behaviors, whereas on the contrary, the second analysis looks for a subspace projecting as well as possible the behavior characterizing the time series. Based on the obtained discriminant spaces, future work will focus on the classification approaches and on the choice of the appropriate proximity measure according to the two above analysis.

REFERENCES

- Banet T.A. and Lebart L. (1984), Local and partial principal component analysis and correspondence analysis. In: COMPSTAT, Proceedings in Computational Statistics, Physica Verlag, Vienna, 113-118.
- Benali H. and Escofier B. (1989) Smooth factorial analysis and factorial analysis of local differences. Multiway data analysis, 327-339, North-Holland, Amsterdam.
- Lebart L. (1969), L'analyse statistique de la contiguïté. Publication de l'I.S.U.P, XVIII, 81-112.
- Le Foll Y. (1982), Pondération des distances en analyse factorielle. Statistiques et Analyse des données, 7, 13-31.
- Mom A. (1988), Méthodologie statistique de la classification de réseaux de transport. Thèse, U.S.T.L., Montpellier.
- Thioulouse J., Chessel D. and Champely S. (1995), Multivariate analysis of spatial patterns: a unified approach to local and global structures. Environmental and Ecological Statistics, 2, 1-14.