The Dual Dynamic Factor Analysis Models *

R. Coppi P. D'Urso

Department of Statistics, Probability and Applied Statistics
University of Rome "La Sapienza", Rome, Italy

Abstract: An exploratory methodology for analyzing three-way arrays of the type "Units $\times$Variables $\times$Times" is described. This is based on the decomposition of the total variation in the array into three components and their modelling by means of a joint utilization of regression and principal components analyses. An extension of this methodology, denominated "dual" approach, is developed, based on interchanging the roles of units w.r.t. variables and times w.r.t. units and variables. An illustration is provided, with reference to the analysis of data concerning the problem of deforestation in Latin America.

1 Introduction

Dynamic Factor Analysis (DFA), originally introduced by Coppi and Zannella (1978) and further developed by Coppi et al. (1996) and Corazziari (1999a, 1999b), is an exploratory methodology for analyzing three-way arrays of the type "Units $\times$Variables $\times$Times", where the variables are quantitative. Let us denote by x_{ijt} the generic element of this array related to unit i $(i = 1, \ldots, I)$, variable j $(j = 1, \ldots, J)$, time t $(t = 1, \ldots, T)$. In its three-dimensional form this array is represented by $\mathbf{X}(I, J, T)$. However, two of the indices may be collapsed into one dimension in 3 different ways. The original proposal by Coppi and Zannella (1978) is based on pooling together units and times along the rows, while associating the variables with the columns of a two-way array. This can be done in two different ways according to whether T is nested within I, thus defining the array $\mathbf{X}(IT, J)$ where t is an index of groups of units, or vice versa I is nested within T, thus defining the array $\mathbf{X}(TI, J)$ where i becomes an index of groups of times. The DFA approach (and related models) referred to $\mathbf{X}(IT, J)$ or $\mathbf{X}(TI, J)$ is called "direct". Following this line of reasoning we can define two new DFA approaches. The first one, referred to the structures $\mathbf{X}(JT, I)$ or $\mathbf{X}(TJ, I)$, may be defined "dual", emphasizing the exchange of roles between units and variables. The second one, concerning the structures $\mathbf{X}(IJ, T)$ or $\mathbf{X}(JI, T)$, may be denominated "tridual", pointing out its "twofold duality" with respect to each of the previous approaches, as will be explained later on. In the following Section we shall give a general description of the methodological bases of DFA, including the direct approach. The dual perspective in DFA is examined in Sec. 3, while Sec. 4 is devoted to an empirical illustration concerning the problem of deforestation in Latin America.

*Invited lecture

W. Gaul et al. (eds.), *Classification, Automation, and New Media*

2 Methodological bases of Dynamic Factor Analysis

Before applying any DFA model, the raw data x_{ijt} are normalized in the following way:

$$z_{ijt} = \frac{x_{ijt}}{\bar{x}_{.j.}}, \quad i = 1, \ldots, I; \quad j = 1, \ldots, J; \quad t = 1, \ldots, T \tag{1}$$

where $\bar{x}_{.j.}$ denotes the overall mean of variable j. Different normalizations might be considered (e.g. dividing each variable by the overall median or standard deviation). Transformation (1) seems to perform well in most situations, eliminating the effects due to different units of measurement and sizes of the characters while preserving the information related to variation, which is a key concept in DFA. In what follows we shall refer to the normalized arrays $\mathbf{Z}(I, J, T)$, $\mathbf{Z}(IT, J)$ and so forth. The general methodological bases of DFA, common to all approaches and models, are:

1. The decomposition of the overall variation of $\mathbf{Z}(I, J, T)$ into three components;

2. The modelisation and analysis of these components by means of a joint utilization of Singular Value Decomposition (SVD) and Regression analysis with respect to time (when applicable).

We shall first illustrate this methodology in connection with the direct approach. In this case the overall variation of $\mathbf{Z}(I, J, T)$ refers to the collapsed arrays $\mathbf{Z}(IT, J)$ or $\mathbf{Z}(TI, J)$, for which the operator $\mathbf{S} = \{s_{jj'}\}, \ j, j' = 1, \ldots, J$, is defined, where

$$s_{jj'} = \frac{1}{IT} \sum_{i,t} \left(z_{ijt} - \bar{z}_{.j.}\right) \left(z_{ij't} - \bar{z}_{.j'.}\right)$$

is the overall covariance between variables j and j'.
The basic decomposition of $\mathbf{S}$ is as follows:

$$\mathbf{S} = {}^{*}\mathbf{S}_I + {}^{*}\mathbf{S}_T + \mathbf{S}_{IT} \tag{2}$$

where the generic elements of ${}^{*}\mathbf{S}_I$, ${}^{*}\mathbf{S}_T$ and $\mathbf{S}_{IT}$ are respectively:

$$\overset{*}{{}_I}s_{jj'} = \frac{1}{I} \sum_{i} \left(\bar{z}_{ij.} - \bar{z}_{.j.}\right) \left(z_{ij'.} - \bar{z}_{.j'.}\right)$$

$$\overset{*}{{}_T}s_{jj'} = \frac{1}{T} \sum_{t} \left(\bar{z}_{.jt} - \bar{z}_{.j.}\right) \left(\bar{z}_{.j't} - \bar{z}_{.j'.}\right)$$

$$\overset{*}{}_{IT}s_{jj'} = \frac{1}{IT}\sum_{i,t}\left(z_{ijt} - \bar{z}_{ij.} - \bar{z}_{.jt} + \bar{z}_{.j.}\right)\left(z_{ij't} - \bar{z}_{ij'.} - \bar{z}_{.j't} + \bar{z}_{.j'.}\right)$$

${}^*\mathbf{S}_I$ describes the "synthetic structure" of the units, independently of time; ${}^*\mathbf{S}_T$ measures the variation due to the "average time evolution" of the variables (for the whole set of units) and, finally, $\mathbf{S}_{IT}$ assesses the residual variation due to the "differential time evolution" of the units (interaction between units and times). With reference to $\mathbf{Z}(IT, J)$ and $\mathbf{Z}(TI, J)$ respectively, we establish the two following decompositions:

$$\mathbf{S} = \bar{\mathbf{S}}_T + {}^*\mathbf{S}_T \tag{3}$$

$$\mathbf{S} = \bar{\mathbf{S}}_I + {}^*\mathbf{S}_I \tag{4}$$

where

$$\bar{\mathbf{S}}_T = {}^*\mathbf{S}_I + \mathbf{S}_{IT}$$

measures the "global structure" of the units (including their differential time evolution) and

$$\bar{\mathbf{S}}_I = {}^*\mathbf{S}_T + \mathbf{S}_{IT}$$

measure the "global dynamics" of the system (average evolution of variables+differential evolution of units).
Two basic DFA models can now be defined.

Model 1 (direct approach)
This is based on decomposition (3). Notice that

$$\bar{\mathbf{S}}_T = \frac{1}{T}\sum_t \mathbf{S}(t) \tag{5}$$

where $\mathbf{S}(t)$ is the covariance matrix at time t. Thus, expression (3) may be interpreted as a classical decomposition of the total variation in the array into "within times" + "between times" variation. The "within times" variation identifies the "global structure" of the units. The corresponding operator, $\bar{\mathbf{S}}_T$, is analyzed by means of a SVD. This is equivalent to carrying out a Principal Components analysis of T clouds of points in R^{J+1}, where the t-th cloud (along the time axis) represents the I units centered with respect to their barycenter

$$\bar{\mathbf{u}}'_{.t} = (\bar{z}_{.1t}, \ldots, \bar{z}_{.Jt}, t) \tag{6}$$

This analysis provides:

a. Component scores for each unit (along the h-th principal axis): F_{ih}. The scores F_{ih} $(i = 1, \ldots, I)$ allow us to represent the synthetic structure of the units, related to the variation of the vectors

$$(z_{i1.}, \ldots, z_{iJ.}), \;\; t = 1, \ldots, T$$

b. Factorial "trajectories" of each unit, given by the component scores $F_{iht}\,(t = 1, \ldots, T)$, obtained by applying the h-th eigenvector of $\bar{\mathbf{S}}_T$ to vectors $(\bar{z}_{i1t}, \ldots, \bar{z}_{iJt})$, $i = 1, \ldots, I$. These trajectories represent the differential time evolution of the units.

The "between times" variation, measured by ${}^*\mathbf{S}_T$, is analyzed by means of a polynomial Time Regression model for the components of $\bar{\mathbf{u}}_{.t}$ [see (6)]:

$$\bar{z}_{.jt} = b_{0j} + b_{1j}t + \ldots + b_{kj}t^k + e_{jt} \qquad j = 1, \ldots, J \tag{7}$$

where the residuals e_{jt} verify the following conditions:

$$cov\left(e_{jt}, e_{j't'}\right) = \begin{cases} w_j \geq 0, & j = j', t = t' \\ 0, & otherwise \end{cases}$$

The parameters matrix $\mathbf{B} = (\mathbf{b}_1, \ldots, \mathbf{b}_J)$ in model (7) is estimated by Least Squares:

$$\hat{\mathbf{B}} = \left(\mathbf{A}'\mathbf{A}\right)^{-1} \mathbf{A}' \, {}^*\mathbf{Z}_I$$

where ${}^*\mathbf{Z}_I = \{\bar{z}_{.jt}\}, t = 1, \ldots, T, j = 1, \ldots, J$ and $\mathbf{A}$ is the design matrix associated with model (7).This model accounts for the average time evolution of the set of variables. Model 1 of DFA may also be interpreted in terms of an "optimal" fitting of the array $\mathbf{Z} = \left\{z_{ijt}\right\}$. In fact we can write

$$z_{ijt} = (z_{ijt} - \bar{z}_{ij.} - \bar{z}_{.jt} + \bar{z}_{.j.}) + (\bar{z}_{ij.} - \bar{z}_{.j.}) + \bar{z}_{.jt}$$

and look for theoretical values $\hat{z}_{ijt}$ such that

$$\Phi = \sum_{i,j,t} \left(z_{ijt} - \hat{z}_{ijt}\right)^2 = min$$

within a class of possible models. If we assume the form

$$\hat{z}_{ijt} = \sum_h a_{jh} F_{iht} + {}_j f(t)$$

we may write

$$\begin{aligned} \Phi \quad = \quad & \sum_{i,j,t} \left\{ \left[(z_{ijt} - \bar{z}_{ij.} - \bar{z}_{.jt} + \bar{z}_{.j.}) + (\bar{z}_{ij.} - \bar{z}_{.j.}) - \sum_h a_{jh} F_{iht} \right] \right. \\ & \left. + \left[\bar{z}_{.jt} -_j f\left(t\right)\right] \right\}^2 = \min \end{aligned} \tag{8}$$

Since the cross product in (8) can be shown to vanish, the minimum is achieved by: a_{jh} = j-th element of the h-th eigenvector of $\bar{\mathbf{S}}_T$, i.e. $\mathbf{a}_h$; F_{iht} = factorial coordinate of unit i at time t on the h-th axis: ${}_jf(t)$ = time regression estimate of the mean value of variable j at time t. The

SOURCE	REFERENCE (A)	ACCOUNTED (B)	INDICATOR (B:A)
Global structure of units	$tr\bar{\mathbf{S}}_T$	$\sum_h \mathbf{a}'_h\,{}^*\mathbf{S}_I\mathbf{a}_h + tr\hat{\mathbf{S}}_{IT}$	I_T
Synthetic structure of units	$tr\,{}^*\mathbf{S}_I$	$\sum_h \mathbf{a}'_h\,{}^*\mathbf{S}_I\mathbf{a}_h$	${}^*I_I$
Differential evolution of units	$tr\mathbf{S}_{IT}$	$tr\hat{\mathbf{S}}_{IT}$	I_{IT}
Average evolution of variables	$tr\,{}^*\mathbf{S}_T$	$tr\,{}^*\hat{\mathbf{S}}_T$	${}^*I_T$
Total	$tr\mathbf{S}$	$\sum_h \mathbf{a}'_h\,{}^*\mathbf{S}_I\mathbf{a}_h + tr\hat{\mathbf{S}}_{IT} + tr\,{}^*\hat{\mathbf{S}}_T$	I

Table 1: Indicators of Quality of Representation (Model 1 - Direct approach)

quality of representation pertaining to Model 1 can be assessed by means of specific indicators summarized in Table 1. where

$$^*\hat{\mathbf{S}}_T = \frac{1}{T}\left[{}^*\mathbf{Z}'_I\left(\mathbf{I} - \frac{1}{T}\mathbf{1}\mathbf{1}'\right)\mathbf{A}\left(\mathbf{A}'\mathbf{A}\right)^{-1}\mathbf{A}'\,{}^*\mathbf{Z}_I\right]$$

$$\hat{\mathbf{S}}_{IT} = \mathbf{S} - \left(\sum_h \lambda_h \mathbf{a}_h\mathbf{a}'_h + {}^*\hat{\mathbf{S}}_T\right)$$

λ_h are the eigenvalues of $\bar{\mathbf{S}}_T$ and $\sum_h$ is extended over the set of components retained in the analysis. In addition, the usual PCA measures of quality of representation for the individual units (such as the squared cosine) can be adopted in the context of the analysis of their synthetic structure. When looking at their evolutive configuration in terms of factorial trajectories, the quality of representation can be assessed by means of a Minimum Spanning Tree visualization (see, e.g., Jolliffe, 1984), computed in R^{J+1} and mapped onto the selected factorial space. It should also be underlined that by introducing appropriate dissimilarity measures in the set of trajectories, these can undergo suitable cluster analysis procedures in order to determine useful classifications of the units (Coppi and D'Urso, 1999; D'Urso, 2000).

Model 2 (direct approach)

This is based on decomposition (4). Notice that

$$\bar{\mathbf{S}}_I = \frac{1}{I}\sum_i \mathbf{S}(i)$$

where $\mathbf{S}(i)$ is the covariance matrix for unit i, letting t play the role of individuals. In this respect, (4) may be interpreted as a decomposition of the total variation of the array into "within units" and "between units" variation.The "within units" variation, measured by $\bar{\mathbf{S}}_I$, identifies the "global structure" of times corresponding to the variation due to the overall dynamics of the array. In this case the two sources of dynamics, namely the average evolution of variables and the differential evolution of units, are simultaneously analyzed by means of a polynomial time Regression model for each unit and variable:

$$z_{ijt} = b_{0ij} + b_{1ij}t + \ldots + b_{kij}t^k + e_{ijt} \qquad i = 1, \ldots, I,\ j = 1, \ldots, J \quad (9)$$

where

$$cov\left(e_{ijt}, e_{i'j't'}\right) = \begin{cases} w_{ij} \geq 0, & i = i', j = j', t = t' \\ 0, & otherwise \end{cases}$$

Least squares estimation of the parameters in (9) provides a way of studying the time evolution of the units for each single variable. For given k, it can be shown that models (7) are the average (over the units) of models (9), for each variable. The differential dynamics of the units can be analyzed by comparing the equations (9) and the corresponding equations (7) (looking at the differences between the homologous parameters). The "between units" variation, measured by $^*\mathbf{S}_I$, identifies the "synthetic structure" of the units and is analyzed by a SVD of $^*\mathbf{S}_I$. Differently from Model 1, in the present case the factorial approach refers only to the mean values $\bar{z}_{ij.}$ which do not contain information related to time evolution.

3 The "dual" perspective in Dynamic Factor Analysis

We now take into consideration the arrays $\mathbf{Z}(JT, I)$ or $\mathbf{Z}(TJ, I)$, on one side, and the arrays $\mathbf{Z}(IJ, T)$ or $\mathbf{Z}(JI, T)$ on the other side. In the former case we may look at the total "Proximity Operator" between units:

$$\mathbf{P} = \left\{ \frac{1}{JT} \sum_{j,t} \left(z_{ijt} - \bar{z}_{i..}\right) \left(z_{i'jt} - \bar{z}_{i'..}\right) \right\}_{i,i' \in I} \qquad (10)$$

In the latter case we may focus our interest on the total "Proximity Operator" between times:

$$\mathbf{Q} = \left\{ \frac{1}{IJ} \sum_{i,j} \left(z_{ijt} - \bar{z}_{..t}\right) \left(z_{ijt'} - \bar{z}_{..t'}\right) \right\}_{t,t' \in T} \qquad (11)$$

Operators (10) and (11) match the previously considered matrix $\mathbf{S}$, which can be interpreted as a "Proximity Operator" between variables. It should be noticed that all quantities in (10) and (11) bear a statistical meaning, since the data z_{ijt} are normalized according to (1) and can be dealt with as "index numbers". Comparing $\mathbf{P}$ with $\mathbf{S}$ we realize that the role of units and variables is permuted. Dual DFA is defined by the application of the DFA machinery to **P**, instead of **S**. When considering **Q**, the "time dimension" looses its characteristic of being an ordered set and is treated in the same way as I and J respectively in **P** and **Q**. The reference arrays of operator **Q**, namely $\mathbf{Z}(IJ, T)$ and $\mathbf{Z}(JI, T)$, differ from the previous ones in that their rows are not related to ordered sets (as it happens, on the contrary, for the set T in the direct and dual approaches). As a consequence, the time regression component of DFA is meaningless in this context. Nonetheless, it makes sense to apply the factorial component of the DFA machinery when considering **Q**. This defines what we may call the "tridual" DFA. We start our illustration by examining in some detail the dual DFA, whereas we shall limit ourselves to giving just a hint to the tridual approach.

Dual Dynamic Factor Analysis

It can easily been shown that the basic decomposition of type (2) holds true also for **P**:

$$\mathbf{P} = {}^*\mathbf{P}_J + {}^*\mathbf{P}_T + \mathbf{P}_{JT} \tag{12}$$

where

$${}^*\mathbf{P}_J = \left\{ \frac{1}{J} \sum_j \left(\bar{z}_{ij.} - \bar{z}_{i..}\right) \left(z_{i'j.} - \bar{z}_{i'..}\right) \right\}_{i,i' \in I} \tag{13}$$

measures the "synthetic structure" of the variables, independently of time,

$${}^*\mathbf{P}_T = \left\{ \frac{1}{T} \sum_t \left(\bar{z}_{i.t} - \bar{z}_{i..}\right) \left(z_{i'.t} - \bar{z}_{i'..}\right) \right\}_{i,i' \in I} \tag{14}$$

represents the variation due to the "average time evolution" of the units (over the whole set of variables), and

$$\begin{aligned}\mathbf{P}_{JT} \quad = \quad & \left\{ \frac{1}{JT} \sum_{j,t} \left(z_{ijt} - \bar{z}_{ij.} - \bar{z}_{i.t} + \bar{z}_{i..}\right) \times \right. \\ & \left. \times \left(z_{i'jt} - \bar{z}_{i'j.} - \bar{z}_{i'.t} + \bar{z}_{i'..}\right) \right\}_{i,i' \in I}\end{aligned} \tag{15}$$

is a measure of the "residual" variation due to the "differential evolution" of the variables (interaction between variables and times). The two decompositions, matching (3) and (4) of the direct approach, are as follows:

$$\mathbf{P} = \bar{\mathbf{P}}_T + {}^*\mathbf{P}_T \tag{16}$$

$$\mathbf{P} = \bar{\mathbf{P}}_J + {}^*\mathbf{P}_J \tag{17}$$

where

$$\bar{\mathbf{P}}_T = {}^*\mathbf{P}_J + \mathbf{P}_{JT} \tag{18}$$

measures the "global structure" of the units (including their differential time evolution) and

$$\bar{\mathbf{P}}_J = {}^*\mathbf{P}_T + \mathbf{P}_{JT} \tag{19}$$

measures the "global dynamics" of the system (given, in this case, by the sum of the average evolution of units and the differential evolution of variables, as compared with the corresponding matrix $\bar{\mathbf{S}}_I$ of the direct approach). In connection with decompositions (18) and (19) respectively, the following two dual DFA models can be set up.

Model 1 (dual approach)
This is based on decomposition (18). In this respect we get

$$\bar{\mathbf{P}}_T = \frac{1}{T}\sum_t \mathbf{P}(t) \tag{20}$$

where $\mathbf{P}(t)$ is the proximity matrix between units at time t. In this case the "within+between times" interpretation of (18) refers to the variation of units instead of variables (as in the direct approach). The "within times" variation of units identifies the "global structure" of variables and is studied by means of a SVD, providing:

a. Component scores for each variable (along the h-th principal axis): F_{hj}. These scores allow us to represent the synthetic structure of the variables, related to the variation of the vectors

$$\left(\bar{z}_{1j.}, \ldots, \bar{z}_{Ij.}\right), \quad j = 1, \ldots, J$$

b. Factorial "trajectories" of each variable, given by the component scores $F_{hjt}(t = 1, \ldots, T)$, obtained by applying the h-th eigenvector of $\bar{\mathbf{P}}_T$ to vectors $(z_{1jt}, \ldots, z_{Ijt})$, $t = 1, \ldots, T$. These trajectories represent the differential time evolution of the variables.

The "between times" variation, measured by ${}^*\mathbf{P}_T$, is analyzed by means of a Time Regression of the following type:

$$\bar{z}_{i.t} = b_{0i} + b_{1i}t + \ldots + b_{ki}t^k + e_{it} \qquad i = 1, \ldots, I \tag{21}$$

with the conditions

$$cov\left(e_{it}, e_{i't'}\right) = \begin{cases} w_i \geq 0, & i = i', t = t' \\ 0, & otherwise \end{cases} \tag{22}$$

Model (21) represents the average time evolution of the units and is estimated by Least Squares, in the same way as illustrated for the corresponding model (7) in the direct approach. The same analogy exists concerning the "reconstruction" of the data z_{ijt} on the basis of Model 1 in the dual version, and the possibility of setting up a similar system of indicators for assessing the quality of representation.

Model 2 (dual approach)
We refer now to decomposition (19), where

$$\bar{\mathbf{P}}_J = \frac{1}{J}\sum_j \mathbf{P}(j) \tag{23}$$

where $\mathbf{P}(j)$ is the proximity matrix between units calculated for variable j with reference to the values observed over the entire period. Therefore, in this case, we get a decomposition of the overall variation into "within variables+between variables" variation. The former component is measured by (23) and identifies the "global dynamics" of the system from the viewpoint of proximity between units (instead of proximity between variables, as in the direct approach). The latter component, ${}^*\mathbf{P}_J$, characterizes the "synthetic structure" of variables, in terms of proximity between units, independently of time. Model 2, in the dual version, utilizes a Time Regression model for $\bar{\mathbf{P}}_J$, which coincides with model (9) of the direct approach. Thus we are using the same model for explaining the global dynamics of the system, looking on one side at the proximity between variables and, on the other side, at the proximity between units. The "synthetic structure" of the variables is analyzed by means of a SVD of $\bar{\mathbf{P}}_J$. Also Model 2, in the dual version, lends itself to an interpretation in terms of data reconstruction. Moreover, the usual indicators of quality of representation can be computed.

Relationships between direct and dual DFA
The above illustration of the dual approach makes it clear that interchanging the roles of units and variables, passing from the direct to the dual DFA models, provides a useful integration as to the information embodied in the array $\mathbf{Z}(I, J, T)$. Looking for instance at Model 1, the direct approach allows us to compute component scores and factorial trajectories for each unit, whereas the same information concerning the variables is only given by the dual approach. It should be underlined, in this connection, that there does not exist an explicit algebraic relationship between the SVD's of matrices $\bar{\mathbf{S}}_T$ and $\bar{\mathbf{P}}_T$. In fact, considering expressions (5) and (20), we notice that

$$\begin{aligned}
\mathbf{S}(t) &= \mathbf{Z}_t'\mathbf{Z}_t - I\left({}_J\bar{\mathbf{z}}_t\, {}_J\bar{\mathbf{z}}_t'\right) \\
\mathbf{P}(t) &= \mathbf{Z}_t\mathbf{Z}_t' - J\left({}_I\bar{\mathbf{z}}_t\, {}_I\bar{\mathbf{z}}_t'\right)
\end{aligned}$$

where ${}_J\bar{\mathbf{z}}_t$ and ${}_I\bar{\mathbf{z}}_t$ are, respectively, the column and row centers of $\mathbf{Z}_t$. Therefore, due to different centerings of the component matrices, the eigenvectors of $\bar{\mathbf{S}}_T$ and $\bar{\mathbf{P}}_T$ do not necessarily define the same latent dimensions (as it happens in classical PCA). However, empirical studies confirm that the principal axes drawn from the two matrices have similar interpretations. This is witnessed, for instance, by the fact that the correlations between the variables and the component scores of the units in the direct approach, show the same pattern as the component scores of the variables in the dual approach.

Tridual Dynamic Factor Analysis
We give just a hint to this case, emphasizing the specificity due to the treatment of Times as columns of the array $\mathbf{Z}(IJ,T)$, which prevents us from using a regression approach in this framework. In any case, using a notation already introduced in the direct and dual models, we get the following decompositions:

$$\begin{aligned} \mathbf{Q} &= {}^{*}\mathbf{Q}_I + {}^{*}\mathbf{Q}_J + \mathbf{Q}_{IJ} && (24)\\ &= \bar{\mathbf{Q}}_I + {}^{*}\mathbf{Q}_I && (25)\\ &= \bar{\mathbf{Q}}_J + {}^{*}\mathbf{Q}_J && (26) \end{aligned}$$

All of these operators can be analyzed by means of SVD providing, in particular, additional information on the role of time with respect to the units and the variables. For example, the SVD of $\bar{\mathbf{Q}}_I$ gives, among other things, the component scores of the times (w.r.t. the whole set of units) and their differential scores across the different units. In this way, we obtain a dual perspective as compared to the component scores and factorial trajectories of the units provided by the SVD of $\bar{\mathbf{S}}_T$ in Model 1 (direct approach). The two perspectives can be matched in order to cast more light on the complex relationship between units and times. Similar considerations can be made for the SVD of the remaining $\mathbf{Q}$-operators (in particular, concerning the duality w.r.t. the dual models). Summing up, it is suggested that the integration of the direct DFA approach with the dual and tridual models greatly improves the capability of the method in drawing information from the original 3-way array and opens up the way to the formulation of appropriate strategies of analysis in this framework.

4 Application

We illustrate, in an extremely concise way, an application of DFA to a 3-way array which has been set up within a study of deforestation in Latin America (see Corazziari, 1999b). The data refer to 19 Countries in Latin America; 27 variables concerning the proportion of different types of land use, the population densities w.r.t. areas differently characterized from the viewpoint of land use, wood production and commerce, gross domestic product (GDP); 24 years (1961, 1970-1992). A specific software for DFA, written in Xlisp-Stat, has been utilized (Corazziari, 1999b). The data have been normalized according to (1). We limit our comments here to the application of Model 1, in both the direct and the dual versions. In both cases a three-dimensional solution has been chosen for the factorial part of the models. The global Quality Indicators for the two models are reported in Table 2: A common interpretation of

SOURCE	DIRECT	DUAL
Global Structure	0.74	0.76
Synthetic Structure	0.72	0.73
Differential Dynamics	0.89	0.83
Average Evolution	0.79	0.80
Total	0.74	0.76

Table 2: Global Quality Indicators for Model 1 (direct and dual versions)

the factorial structure in the two models is possible, taking into account the correlations of the variables (units) with the component scores of the units (variables) as previously underlined. In fact the first factorial dimension, accounting for nearly 40% of the variation due to the global structure in both models, is connected with the degree and type of agricultural exploitation of the land, while the second dimension ($\simeq$20%) is linked with the import/export of industrial wood and GDP. The third dimension ($\simeq$15%) enhances the opposition between the extension of forests and the extension of lands devoted to permanent pasture. Specific groups of Countries (direct analysis) and of variables (dual analysis) can be singled out on the three principal planes. A few Countries lie apart from the others, showing specific features on the various factorial axes, namely Belize, Chile, Guyana, Surinam and Uruguay. Similarly, a few variables seem to have the major impact on the global structure: Forest surface, Total Agricultural surface, Permanent Pasture surface, Farmers density, Production and Import of Industrial Wood, GDP. These results are also reflected by the clusters obtained from the analysis of trajectories of the units and variables. As to the time regression analyses of the operators $^*\mathbf{S}_T$ and $^*\mathbf{P}_T$, it can be observed that for most of the variables the average evolution is linear but for the production of fuel and coal wood, which show a cyclical component acting together with an increasing trend. A linear trend is also suitable for describing the average evolution of the Countries, although a more detailed analysis for the group of variables related to wood production and commerce and to GDP points out nonlinear trends for many Countries. We shall not dwell, in this paper, on the other results of the analysis, as well as on the detailed examination of the features emphasized in the previous brief comment. We just remark the usefulness of DFA, in its extended form, as a sensitive filter capable to translate the complexity of the data into systematic information concerning the structure and the dynamics of the array $\mathbf{Z}(I, J, T)$.

References

COPPI, R.; BLANCO, J.; CAMAÑO, G. and CORAZZIARI, I. (1996): Descomposicin Factorial y Regresiva de la Variabilidad de un Array a tres Vias. In Seminario de Capacitación y Investigación, Bogotà, Colombia, 355-392. Also published in Quantum, vol. 4, 10, 1999, Montevideo, Uruguay, 81-107.

COPPI, R. and D'URSO, P. (1999): The Geometric Approach to the Comparison of Multivariate Time Trajectories. *Proceedings of CLADAG 99, Rome, Italy, 177-180.*

COPPI, R. and ZANNELLA, F. (1978): L'Analisi Fattoriale di una Serie Temporale Multipla relativa allo stesso Insieme di Unità Statistiche. *Proceedings of the XXIX Scientific Meeting of the Italian Statistical Society (SIS), Bologna, Italy, 61-77.*

CORAZZIARI, I. (1999a): Dynamic Factor Analysis. In: Vichi M. and Opitz O. (eds.): Classification and Data Analysis. Theory and Application, Springer-Verlag, Heidelberg, 171-178.

CORAZZIARI, I. (1999b): Analisi di dati Longitudinali Multivariati. Unpublished doctoral dissertation, University of Naples "Federico II", Naples, Italy.

D'URSO, P. (2000): Classificazione Fuzzy per Matrici a Tre Vie Temporali. Unpublished doctoral dissertation, University of Rome "La Sapienza", Rome, Italy.

JOLLIFFE, I.T. (1985): Principal Component Analysis. Springer-Verlag, New York.