

Time Series components separation based on Singular Spectral Analysis visualization: an HJ-biplot method application

Alberto Oliveira da Silva^{1,2,*}, Adelaide Freitas^{1,2}

¹*Department of Mathematics, University of Aveiro, Portugal*

²*Center for Research & Development in Mathematics and Applications (CIDMA), University of Aveiro, Portugal*

Abstract The extraction of essential features of any real-valued time series is crucial for exploring, modeling and producing, for example, forecasts. Taking advantage of the representation of a time series data by its trajectory matrix of Hankel constructed using Singular Spectrum Analysis, as well as of its decomposition through Principal Component Analysis via Partial Least Squares, we implement a graphical display employing the biplot methodology. A diversity of types of biplots can be constructed depending on the two matrices considered in the factorization of the trajectory matrix. In this work, we discuss the called HJ-biplot which yields a simultaneous representation of both rows and columns of the matrix with maximum quality. Interpretation of this type of biplot on Hankel related trajectory matrices is discussed from a real-world data set.

Keywords Singular Spectrum Analysis, NIPALS algorithm, Biplots

AMS 2010 subject classifications 62-09, 62M15

DOI: 10.19139/soic-2310-5070-897

1. Introduction

Time series (TS) data emerge in many scientific fields. The analysis of this type of data can be based on the time or the frequency domain. When its evaluation is based on time-domain, parametric models are proposed. When the study is in terms of frequency-domain, approaches using non-parametric models, such as the Spectral Analysis, are usually developed. A TS can be classified as stationary if properties of the underlying generation process do not change over time, e.g., the mean and the variance are constants. Otherwise, the TS is said to be nonstationary and, in this case, can reveal a trend, i.e., a smooth and slowly varying part of the series. Any TS can be decomposed into a variety of components, for instance, (1) global trend (only for nonstationary TS), (2) oscillatory shape, e.g., a seasonal variation, and (3) irregular component, or noise. Generally, it is possible to model the trend by mean of a low degree polynomial function

$$f(t) = a_0 + a_1t + \dots + a_mt^m$$

where t is the time, a_i are real constants and m denoting here the degree of the polynomial function. Likewise, the regular oscillatory, with a period p , can be synthesized as a function $g(t)$ defined by a linear combination of sines and cosines with constant coefficients, i.e., expressed as a Fourier series

$$g(t) = \sum_{m=0}^p (\alpha_m \cos(\omega_m t) + \beta_m \sin(\omega_m t))$$

*Correspondence to: Alberto Oliveira da Silva (Email: albertos@ua.pt). Center for Research & Development in Mathematics and Applications (CIDMA), University of Aveiro. 3810-193 Aveiro, Portugal.

where $\omega_m = 2\pi m/p$. Thus, a classic model of a TS may be obtained by a systematic component, given by the addition (or multiplication) of $f(t)$ and $g(t)$, plus a random component defined by a white noise, $\{\varepsilon_t\}$, which is independent of the signal.

Singular Spectrum Analysis (SSA) is a methodology used in Time Series Analysis for many different purposes, such as exploratory inspection and forecasting [6]. Differently from other techniques, the SSA does not take into consideration whether the model is additive or multiplicative. Besides, no trend model or previous knowledge about the periodicity of the series is required to apply SSA [5]. In the basic version of SSA, the object is a one-dimensional real-valued TS and consists of two successive stages. In the first one, the decomposition stage, the TS is transformed into a Hankel matrix, named as trajectory matrix, on which the Singular Value Decomposition (SVD) is applied, resulting in the summation of rank-one matrices. Next, in the reconstruction stage, some of these rank-one matrices are grouped appropriately (grouping step). From those groups, in the so-called diagonal averaging step, an approximation for the original object or components of the TS, like trend, oscillatory shape, and noise, separately, can be obtained. The form of a TS and the eigenvectors resulting from the decomposition of its trajectory matrix are related. Thus, the graphical representation of these eigenvectors is a proper way to visualize the components of a TS. For example, a plot of the first eigenvector is suitable to evaluate the existence of a trend. In turn, a scatterplot of two eigenvectors close to each other can determine a geometric pattern in some cases, being useful for assessing the existence of a seasonal component ([6] for more details). This work presents a graphical tool based on pairs of eigenvectors that, in the same plot, combines more information and can lead to the identification of relevant features of a TS.

For any matrix \mathbf{Z} whose rank is r , the SVD factorization provides the best approximation matrix $\tilde{\mathbf{Z}}$, in the least-squares sense, whose rank is less than r . Further, if $\tilde{\mathbf{Z}}$ has rank 2, then the SVD allows practical graphical representations of both rows and columns of the approximation matrix employing the Biplot method [2, 3]. Biplots provide easier interpretations and are much more informative than the traditional scatterplots, beyond that might facilitate the work in the grouping step in SSA. Several types of biplots can be constructed depending on how the three factors identified by SVD are aggregated to obtain only two factors. Herein, the option is the HJ-biplot method proposed by Galindo (1986), which yields a simultaneous representation of both rows and columns of $\tilde{\mathbf{Z}}$ with maximum quality [3].

Taking the Hankel-related trajectory matrix arisen from Basic SSA, an HJ-biplot type exploratory tool is constructed to visualize and identify patterns in nonstationary TS, and its interpretation is discussed. This work suggests the factorization of the trajectory matrix using the Nonlinear Iterative Partial Least Squares (NIPALS) algorithm [14] since it is capable of dealing with missing values, commonly in TS, without employing any imputation method [13, 15]. For complete matrices, it is essential to point out that the NIPALS algorithm provides equivalent results to their factorization via SVD concerning the singular vectors and the singular values.

The paper is organized as follows. In Section 2, a short overview of theoretical background related to methods involved in this work is provided. In Section 3, the proposed biplot approach to the SSA method and its interpretation are discussed. In Section 4, the suggested technique is performed on a real-world TS using the statistical software R [12]. Some R-code fragments are indicated. Conclusions are presented in Section 5.

2. Methods

2.1. Basic Singular Spectrum Analysis

Consider a real-valued TS $Y = (y_1, \dots, y_n)$ of length n . Basic SSA is a model-free tool used to recognize and identify the structure of Y [6]. As aforementioned above, the SSA consists of two complementary stages: decomposition and reconstruction. Each stage in this algorithm includes two steps.

First Stage: Decomposition.

Let ℓ ($1 < \ell < n$) an integer value representing the so-called window length, as well as $\kappa = n - \ell + 1$. Hereupon, the embedding procedure consists in representing Y in κ lagged vectors $\mathbf{x}_1, \dots, \mathbf{x}_\kappa$, each one of size ℓ (ℓ -lagged vectors), i.e., $\mathbf{x}_j = [y_j, \dots, y_{j+\ell-1}]'$, $1 \leq j \leq \kappa$. This sequence of κ vectors forms the trajectory matrix $\mathbf{X} = [\mathbf{x}_1 \cdots \mathbf{x}_\kappa]$, a Hankel matrix that has in its columns the ℓ -lagged vectors. Thus, the trajectory matrix consists

of the transformation of a time series into a Hankel matrix by means of an embedding operator \mathcal{T} , such that

$$\mathcal{T}(Y) = \mathbf{X} = \begin{bmatrix} y_1 & y_2 & y_3 & \cdots & y_\kappa \\ y_2 & y_3 & y_4 & \cdots & y_{\kappa+1} \\ y_3 & y_4 & y_5 & \cdots & y_{\kappa+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_\ell & y_{\ell+1} & y_{\ell+2} & \cdots & y_n \end{bmatrix}$$

Next, SVD is executed for the trajectory matrix \mathbf{X} resulting

$$\mathbf{X} = \sum_{i=1}^d \sqrt{\lambda_i} \mathbf{u}_i \mathbf{v}_i',$$

where $d = \text{rank}(\mathbf{X})$, λ_i , $i = 1, \dots, d$, are the eigenvalues of the matrix $\mathbf{X}'\mathbf{X}$ arranged in decreasing order of magnitudes ($\lambda_i > 0$), and associated to the orthonormal system of the eigenvectors $\{\mathbf{v}_1, \dots, \mathbf{v}_d\}$ of $\mathbf{X}'\mathbf{X}$, and

$$\mathbf{u}_i = \mathbf{X}\mathbf{v}_i / \sqrt{\lambda_i}.$$

The elements of the triple $(\sqrt{\lambda_i}, \mathbf{u}_i, \mathbf{v}_i)$ are also known as singular values, left and right singular vectors of \mathbf{X} , respectively. Besides, defining

$$\mathbf{X}_i = \sqrt{\lambda_i} \mathbf{u}_i \mathbf{v}_i',$$

\mathbf{X} can be represented by a sum of d 1-rank matrices, i.e.,

$$\mathbf{X} = \mathbf{X}_1 + \cdots + \mathbf{X}_d. \quad (1)$$

Second Stage: Reconstruction.

Once the expansion (1) has been determined, the second stage of SSA starts with the partitioning of the index set $\{1, \dots, d\}$ into disjoint subsets I_j , $j = 1, \dots, p$, leading to the decomposition given as follows

$$\mathbf{X}_I = \mathbf{X}_{I_1} + \cdots + \cdots + \mathbf{X}_{I_p},$$

where $\mathbf{X}_I = \sum_{i \in I} \mathbf{X}_i$. The intention of the grouping procedure is the separation of the additive components of the TS [7]. The objective of the next phase, the diagonal averaging step, is to take each matrix \mathbf{X}_{I_j} of the grouping step and transform it into a Hankel matrix $\tilde{\mathbf{X}}_{I_j}$, converting the result into a TS [6] by means of

$$\tilde{Y}^{I_j} = \mathcal{T}^{-1}(\tilde{\mathbf{X}}_{I_j}).$$

2.2. PCA through NIPALS

The NIPALS algorithm belongs to the Partial Least Squares (PLS) family, a set of iterative algorithms that implement a wide range of multivariate explanatory and exploratory techniques. The NIPALS is designed as an iterative estimation method for Principal Component Analysis (PCA), that computes the principal components through an iterative sequence of simple ordinary least squares regressions [13, 14]. NIPALS on \mathbf{X} produces a decomposition of the matrix so that the principal components are computed one-by-one [13], providing equivalent results to the SVD concerning the singular vectors and the singular values. A particular feature of the NIPALS algorithm is that, in each iteration, only present data are considered in the regressions performed, ignoring the missing elements. It is equivalent to defining all missing points as zero in the least-squares objective function. Therefore, in the case of missing data, no imputation method is necessary when applying NIPALS, which can be evaluated as an advantage over the SVD.

Considering a d -rank trajectory matrix, the algorithm decomposes \mathbf{X} as a sum of d 1-rank matrices in terms of the outer product of two vectors, a score \mathbf{t}_i and a loading \mathbf{p}_i , so that

$$\mathbf{X} = \mathbf{t}_1\mathbf{p}'_1 + \dots + \mathbf{t}_d\mathbf{p}'_d.$$

The elements of the scores vector \mathbf{t}_i are the projections of the sample points on the principal component (PC) direction, while each loading in \mathbf{p}_i is the cosine of the angle between the component direction vector and a variable axis [4]. The NIPALS first computes \mathbf{t}_1 and \mathbf{p}_1 from \mathbf{X} and, then, the outer product $\mathbf{t}_1\mathbf{p}'_1$ is subtracted from \mathbf{X} to calculate the residual matrix \mathbf{E}_1 . After, \mathbf{E}_1 is used to compute \mathbf{t}_2 and \mathbf{p}_2 , and the residual \mathbf{E}_2 is calculated subtracting $\mathbf{t}_2\mathbf{p}'_2$ from \mathbf{E}_1 , and so on until to obtain \mathbf{t}_d and \mathbf{p}_d . The NIPALS algorithm is shown in Algorithm 1.

Algorithm 1 (NIPALS)

Input: $\mathbf{E}_0 = \mathbf{X}$

Output: $\mathbf{P} = [\mathbf{p}_1 \dots \mathbf{p}_d]$, $\mathbf{T} = [\mathbf{t}_1 \dots \mathbf{t}_d]$

For all $i = 1, \dots, d$ **Do**

1) initialize \mathbf{t}_i

2) **Repeat**

$$\mathbf{p}_i = \mathbf{E}'_{i-1}\mathbf{t}_i/\mathbf{t}'_i\mathbf{t}_i$$

$$\mathbf{p}_i = \mathbf{p}_i/\|\mathbf{p}_i\|$$

$$\mathbf{t}_i = \mathbf{E}_{i-1}\mathbf{p}_i$$

until convergence of \mathbf{p}_i

3) $\mathbf{E}_i = \mathbf{E}_{i-1} - \mathbf{t}_i\mathbf{p}'_i$

End For

From the internal relations in each iteration of the NIPALS algorithm, and after normalizing \mathbf{t}_i , such that

$$\mathbf{t}_i^* = \mathbf{t}_i/\|\mathbf{t}_i\| \iff \mathbf{t}_i = \sqrt{\mathbf{t}'_i\mathbf{t}_i}\mathbf{t}_i^*,$$

the following equations can be verified [13]:

$$\mathbf{E}'_{i-1}\mathbf{E}_{i-1}\mathbf{p}_i = \lambda_i\mathbf{p}_i \tag{2}$$

$$\mathbf{E}_{i-1}\mathbf{E}'_{i-1}\mathbf{t}_i^* = \lambda_i\mathbf{t}_i^*, \tag{3}$$

where $\lambda_i = \mathbf{t}'_i\mathbf{t}_i$ is the eigenvalue of both matrices $\mathbf{E}'_{i-1}\mathbf{E}_{i-1}$ and $\mathbf{E}_{i-1}\mathbf{E}'_{i-1}$, as well as \mathbf{p}_i and \mathbf{t}_i^* are their corresponding eigenvectors with unit norm. In the first iteration of the algorithm, i.e., for $i = 1$, the equations in (2) and (3) are reduced to

$$\mathbf{X}'\mathbf{X}\mathbf{p}_1 = \lambda_1\mathbf{p}_1$$

$$\mathbf{X}\mathbf{X}'\mathbf{t}_1^* = \lambda_1\mathbf{t}_1^*$$

and it is clear that the first normalized score vector and the first loading vector are exactly the first left and right singular vectors of \mathbf{X} , respectively, as well as $\sqrt{\mathbf{t}'_1\mathbf{t}_1}$ returns the first singular value of \mathbf{X} . Moreover, for $i > 1$, and considering the column space of \mathbf{X} , the NIPALS computes the i -th principal component over the orthogonal complement of the subspace $\mathbf{t}_1\mathbf{p}'_1 + \dots + \mathbf{t}_{i-1}\mathbf{p}'_{i-1}$, which is equivalent to the SVD approach of imposing the orthogonality restriction among singular vectors when maximizing its objective function. It implies yet that, $\forall i = 1, \dots, d$, \mathbf{t}_i^* and \mathbf{p}_i are equal to the left and right singular vectors of the SVD of \mathbf{X} , respectively, and each $\sqrt{\mathbf{t}'_i\mathbf{t}_i}$ is the corresponding singular value. The NIPALS decomposition of \mathbf{X} is then given by

$$\mathbf{X} = \sqrt{\mathbf{t}'_1\mathbf{t}_1}\mathbf{t}_1^*\mathbf{p}'_1 + \dots + \sqrt{\mathbf{t}'_d\mathbf{t}_d}\mathbf{t}_d^*\mathbf{p}'_d. \tag{4}$$

Defining the matrix Σ as a diagonal matrix containing the singular values $\sqrt{\mathbf{t}'_i\mathbf{t}_i}$ arranged in decreasing order, one can write the matrix form of the expansion (4) as

$$\mathbf{X} = \mathbf{T}^*\Sigma\mathbf{P}', \tag{5}$$

where \mathbf{T}^* is the (unit-)scores matrix whose column vectors \mathbf{t}_i^* are orthonormal, and \mathbf{P} is the (unit-)loadings matrix whose column vectors \mathbf{p}_i are also orthonormal.

2.3. Biplots

Biplot is a 2- or 3-dimensional graph that allows the joint plotting of both objects and variables of multivariate data sets [2]. Biplots can reveal essential characteristics of multivariate data structure, e.g., patterns of correlations between variables or similarities between observations [8]. Consider a d -rank target ($\ell \times \kappa$) data matrix \mathbf{X} with factorization in the form

$$\mathbf{X} = \mathbf{GH}', \quad (6)$$

where \mathbf{G} is a ($\ell \times q$) matrix and \mathbf{H} is a ($\kappa \times q$) matrix, with $q \leq d$. The matrices \mathbf{G} and \mathbf{H} create two sets of q -dimensional points. If $q = 2$ then the rows and columns of \mathbf{X} can be simultaneously represented in the so called biplot, in which the rows of \mathbf{G} are reproduced by points and the columns of \mathbf{H}' are depicted as vectors connected to the origin (arrows). When $q > 2$, the best 2-rank approximation of \mathbf{X} , in the sense of least square, is considered. Thus, a (2-dimensional) biplot displays both row markers $\mathbf{g}_1, \dots, \mathbf{g}_\ell$ and column markers $\mathbf{h}_1, \dots, \mathbf{h}_\kappa$ of \mathbf{X} , such that the inner product $\mathbf{g}'_i \mathbf{h}_j$ provides an approximation to the element x_{ij} of \mathbf{X} [9]. Based on SVD of \mathbf{X} , many factorization form of \mathbf{X} can be taken into account, and hence different choices of \mathbf{G} and \mathbf{H} , and types of biplots [8]. From the equation (5), and considering $\mathbf{G} = \mathbf{T}^*$ and $\mathbf{H} = \mathbf{P}\Sigma$, the resultant factorization is characterized by preserving the column metrics of \mathbf{X} , and the associated biplot is called Gabriel biplot, or classic biplot or covariance biplot or GH-biplot. In this case, if \mathbf{X} has been centered by columns, this type of biplot satisfies the following properties [9]:

- The norm of a column marker \mathbf{h}_j is proportional to the standard deviation of the respective variable;
- The cosine of the angle formed by column markers approximates the correlation between the related variables;
- The columns are better represented than the rows in terms of quality.

On the other hand, by defining $\mathbf{G} = \mathbf{T}^*\Sigma$ and $\mathbf{H} = \mathbf{P}$, this factorization will preserve the metric of the rows in the so-called form biplot or JK-biplot, in which the Euclidean distances between the row markers approximate the Euclidian distances between the respective individuals in the full space, and the quality of representation of the rows is better than the columns.

Based on SVD, a different type of biplot, called HJ-biplot, was proposed in 1986 by Galindo [3] in which optimal quality representation of the ℓ rows and the κ columns of \mathbf{X} is ensured in the same Euclidean space. An HJ-biplot version based on NIPALS instead of the SVD can be constructed taking the rows of the matrix $\mathbf{J} = \mathbf{T}^*\Sigma$ as row markers, and the rows of the matrix $\mathbf{H} = \mathbf{P}\Sigma$ as column markers of \mathbf{X} . Indeed, from the NIPALS decomposition in (5), results that

$$\mathbf{XP} = \mathbf{T}^*\Sigma\mathbf{P}'\mathbf{P} = \mathbf{T}^*\Sigma.$$

So, the ℓ rows of the matrix $\mathbf{J} = \mathbf{T}^*\Sigma$ correspond to the projections of the ℓ points represented by the rows (individuals) of \mathbf{X} onto the subspace spanned by the loading vectors \mathbf{p}_1 and \mathbf{p}_2 , that is, the best-fit two-dimensional subspace for \mathbf{X} . Likewise, the κ rows of the matrix $\mathbf{H} = \mathbf{P}\Sigma$ coincide with the projections of the κ points expressed by the columns (variables) of \mathbf{X} onto the subspace spanned by the normalized score vectors \mathbf{t}_1^* and \mathbf{t}_2^* , as seen below:

$$(\mathbf{T}^*)'\mathbf{X} = (\mathbf{T}^*)'\mathbf{T}^*\Sigma\mathbf{P}' \iff \mathbf{X}'\mathbf{T}^* = \mathbf{P}\Sigma$$

From (5) follows that both row and column representations of \mathbf{X} , \mathbf{XP} and $\mathbf{X}'\mathbf{T}^*$, respectively, are related since

$$\mathbf{B} = \mathbf{X}'\mathbf{A}\Sigma^{-1} \quad \text{and} \quad \mathbf{A} = \mathbf{XB}\Sigma^{-1},$$

where $\mathbf{B} = \mathbf{X}'\mathbf{T}^*$ and $\mathbf{A} = \mathbf{XP}$. It means that the coordinates of the variables can be expressed as a weighted average of the coordinates of the individuals, and vice-versa. As a consequence, it allows the representation of the rows and columns in the same cartesian coordinates system with optimal quality of representation [3, 9]. However, in the HJ simultaneous representation, the inner product $\mathbf{j}'_i \mathbf{h}_j$ does not provide an approximation to the element x_{ij} , $i = 1, 2, \dots, \ell$ and $j = 1, 2, \dots, \kappa$, of \mathbf{X} anymore, and consequently

$$\mathbf{X} \neq \mathbf{JH}'$$

Accordingly, the interpretation of the HJ-biplot representation can be performed as follows:

- The distance between points corresponds to how different the associated individuals are (dissimilarities), just like in JK-biplots;
- As it occurs in GH-biplot, the size of an arrow (variable) is proportional to the standard deviation of the associated variable, i.e., the longer the arrow, the greater the correspondent standard deviation;
- The cosine of the angle between arrows approximates the correlation between the variables they represent. Thus, if the angle is next to 90 degree it indicates a poor correlation, while an angle close to 0 degree or 180 degree suggests a strong correlation, being positive in the first case and negative in the other.

3. The SSA-HJ-Biplot

In the basic version of the SSA, the trajectory matrix that will be decomposed by the NIPALS algorithm has some peculiarities in relation to the usual multivariate data matrix. Instead of individuals and variables, the rows and columns of the Hankel trajectory matrix represent κ -lagged and ℓ -lagged vectors of a univariate time series, respectively. A row marker, determined by the rows of $\mathbf{T}^* \Sigma$, i.e., $\mathbf{j}'_i = \mathbf{t}'_i$, $i = 1, \dots, \ell$, is depicted as a point in the SSA-HJ-biplot and corresponds to a κ -lagged vector. Each HJ-biplot point may receive a label from identifying the period, e.g., month or year, in which that κ -lagged vector begins, and thus improve the graph interpretation regarding the data. It means that the points in the SSA-HJ-biplot can represent not only the κ -lagged vectors that start in a given period but also the month itself. In turn, an arrow represents the column marker associated with a ℓ -lagged vector. An SSA-HJ-biplot uses any two principal components to visualize information about a TS in an integrative way, since the row and column markers are displayed simultaneously on the same graph, with maximum representation quality. In its turn, each PC is associated with a TS component, e.g., trend, seasonality, and noise, explaining a proportion of the variability of the data, given by

$$PC_{\%}(i) = \frac{\mathbf{t}'_i \mathbf{t}_i}{\sum_{j=1}^d \mathbf{t}'_j \mathbf{t}_j}. \tag{7}$$

Some auxiliary graphs can reveal this relationship between a PC and a TS component. For instance, in a scree plot of $\sqrt{\mathbf{t}'_i \mathbf{t}_i}$ [5], most of the time the first principal components are related to highlighted singular values, indicating an association with the trend. Once these PCs are identified, one can visualize the trend plotting each one of these PCs against an index $j = 1, \dots, \kappa$. Some precautions are necessary to obtain the best results when constructing the SSA-HJ-biplot. For example, the window length ℓ has to be large enough so that each ℓ -lagged vector captures a substantial part of the behavior of the TS [6]. Still, but at the same time, it should permit the interpretability of the graphics display. A window length equals to $n/2$ provides both capabilities because it allows for a most detailed decomposition [6]. Beyond that, it is worth keeping in mind, the higher the percentage of variability explained, the better the quality of the adjustment of the SSA-HJ-biplot [3].

The interpretation of the SSA-HJ-biplot for a Hankel trajectory matrix is performed as follows:

- Proximity of points. Biplot points whose euclidean distances are small imply similarity in the behavior of the associated κ -lagged vectors; hence, if there is a natural number π such that

$$\|\mathbf{g}_t - \mathbf{g}_{t+\pi}\| \approx 0, \quad \forall t,$$

then,

$$\|(y_t, \dots, y_{t+\kappa-1}) - (y_{t+\pi}, \dots, y_{t+\pi+\kappa-1})\| \approx 0, \quad \forall t.$$

This fact leads to suspect that $(y_t, \dots, y_{t+\kappa-1}) = (y_{t+\pi}, \dots, y_{t+\pi+\kappa-1})$, for all t , which means that the TS might have periodic fluctuations (seasonality) with period π ;

- Length of biplot vectors. If some arrows have roughly the same size, this indicates that the correspondent ℓ -lagged vectors have standard deviation also close; hence, if there is a natural number $\tau < \kappa$ such that

$$\|\mathbf{h}_\tau\| \approx \|\mathbf{h}_{\tau+j}\|, \quad \forall j = 1, 2, \dots, \kappa - \tau,$$

then,

$$\text{var}(\mathbf{x}_\tau) \approx \text{var}(\mathbf{x}_{\tau+j}), \quad \forall j = 1, 2, \dots, \kappa - \tau,$$

which suggests that the TS $(y_1, \dots, y_t, \dots, y_n)$ might be generated by a variance stationary process for $t > \tau$;

- Angle between two biplot vectors. If the angle between two arrows is close to 90 degrees, a correlation near to zero is expected which means no similar behaviors of the respective ℓ -lagged vectors. If the angle between two arrows is next to 0 (180, resp.) degree, then a strong and positive (negative, resp.) correlation between the two ℓ -lagged vectors associated is suspected. Thus, if there exists a natural number $\eta < \kappa$ such that

$$|\cos(\angle(\mathbf{h}_j, \mathbf{h}_{j+\eta}))| \approx 1, \quad \forall j = 1, \dots, \tau,$$

for some $\tau \in \{1, \dots, \kappa - \eta\}$, then

$$|\text{cor}(\mathbf{x}_j, \mathbf{x}_{j+\eta})| \approx 1, \quad \forall j = 1, \dots, \tau.$$

Consequently, there are real constants a and $b \neq 0$ such that $y_{j+\eta} \approx a + by_j$, for $j = 1, \dots, \tau$. This means that (y_1, \dots, y_κ) might be generated by a τ -th order stationary process.

As a rule, a singular value represents the contribution of the corresponding PC in the form of the TS. As the trend generally characterizes the shape of a TS, its singular values are higher than the others, that is, they are the first eigenvalues [1]. It means that the directions of the highest variability of a time series are related to the trend, and as mentioned before, it can be modeled employing a low degree polynomial function, such as in (1). Still, when two singular values are close enough, i.e.,

$$\sqrt{\mathbf{t}'_i \mathbf{t}_i} \approx \sqrt{\mathbf{t}'_h \mathbf{t}_h},$$

this is an evidence of the formation of plateaus in the scree plot and indicates that the associated PC is informative about the oscillatory components of the TS [6], as long as the principal component explains high variability of the data. It occurs because the periodical shape of a TS can be expressed as a Fourier series, as in (1). Consequently, for each m , the sine and cosine of $(\omega_m t)$ will determine orthogonal directions of a pair of PC, and the associated singular values will be close to each other. Apart from the interpretation of similarities using Euclidean distances, the projections of the biplot points into a PC axis helps in the identification of the TS components. If the projections evolve in time in the same principal component growth direction, it means this PC is associated with the trend, as well as the trend is crescent. Otherwise, if the evolution in time occurs in the opposite direction, the trend is decreasing. Moreover, this procedure allows detecting a trend change direction quickly. In turn, a pattern in terms of proximity between the projections can occur. In this case, it indicates the correspondence between the PC and the periodicity of the TS.

4. Example

In this section, a real-world TS is used to demonstrate the capabilities of the SSA-HJ-biplot. Two R-libraries accompany this study: `imputeTS` [10] which contains a function for obtaining graphical representations of TS with missing data, and `nipals` [16] which is aimed to perform PCA using NIPALS algorithm. This TS (data in the R-code below) contains the records of the carbon dioxide concentration in the Earth's atmosphere, measured monthly from January of 1965 to December of 1980 ($n = 192$) at an observing station on Mauna Loa in Hawaii [11]. This time series is referred to as TS CO2 in this work and presents missing data, as can be verified in Figure 1. For constructing this plot, the following R-code was used:

```
> Y <- ts(data, start=1965, end=1980, frequency = 12)
> library(imputeTS)
> plotNA.distribution(Y, main="Monthly Dioxide Carbon Concentration
+ (ts CO2)", xlab="", ylab="CO2 concentration")
```

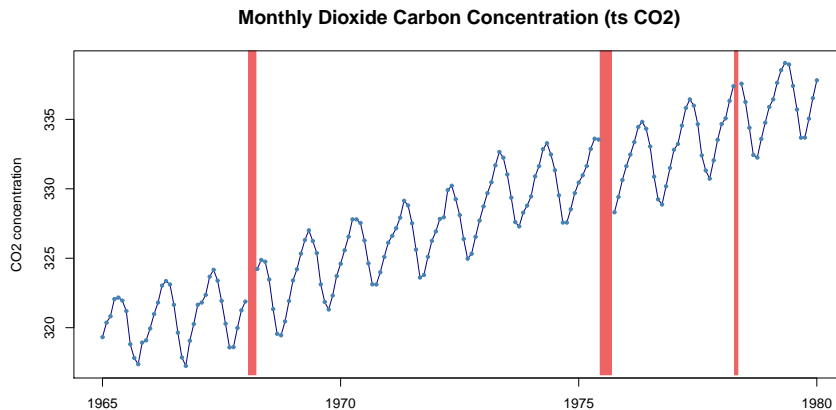



Figure 1. Records of carbon dioxide (CO₂) concentration in the Earth's atmosphere measured monthly from January of 1965 to December of 1980 at an observing station on Mauna Loa in Hawaii. Missing data are represented in red.

As mentioned before, the NIPALS algorithm handles the missing values conveniently without the need to complete the data. For the execution of the NIPALS algorithm on the $(\ell \times \kappa)$ trajectory matrix (X), the following R-code was used:

```
> n = 192; L = ceiling(n/2); K = n - L + 1
> X = outer((1:L), (1:K), function(x,y) data[(x+y-1)])
> library(nipals)
> res = nipals(X, center = TRUE, scale = FALSE)
```

In the embedding step of the SSA, the window length used was $\ell = n/2 = 96$ observations, resulting in $\kappa = 97$. The R-object `res` above contains, among others, the (unit-)scores matrix \mathbf{T}^* , the diagonal matrix $\mathbf{\Sigma}$, and the (unit-)loadings matrix \mathbf{P} related to the NIPALS decomposition (5) of the trajectory matrix, and are computed as:

```
> Tstar = res$scores
> Sigma = diag(res$eig)
> P = res$loadings
```

respectively. Table 1 shows the proportion of variability explained by the ten first PC and calculated according to (7). Therefore, it turns out that the five first PC explain about 98% of the data variability, with less than 2% remaining from the 6th PC onwards. In its turn, Figure 2 brings the scree plot, in which the dominant singular value $\sqrt{\mathbf{t}'_1 \mathbf{t}_1}$ represents the 1st PC and explains about 67% of the data variability, being associated with the trend. Figure 3 shows the first SSA-HJ-biplot, in which the biplot points are labeled with the month and year when the κ -lagged vector starts, ranging from January of the first year (J1) to December of the eighth year (D8). The biplot markers displayed in Figure 3 were obtained using the following R-code:

```
> J = Tstar %*% Sigma
> H = P %*% Sigma
> points(J[,1], J[,2], col="blue")
> arrows(0,0,H[,1],H[,2],length=0.1,lwd=1, col="grey")
```


Table 1. Proportion of variability explained by the ten first principal components

PC (i)	Variance (%)	PC (i)	Variance (%)
1	66.830	6	0.376
2	14.707	7	0.316
3	14.444	8	0.291
4	1.096	9	0.106
5	1.070	10	0.080

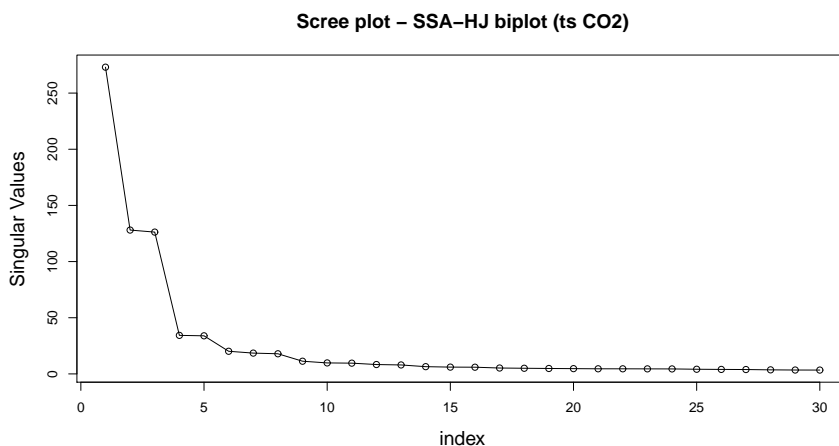


Figure 2. Scree plot of the singular values of the trajectory matrix after NIPALS decomposition.

```
> arrows(0,0,H[1:6,1],H[1:6,2],length=0.1,lwd=1,col="black")
```

It can be verified in Figure 3 that as the 1st PC increases, the projection of the points representing the κ -lagged vectors also progress over time, indicating a crescent trend. On the other hand, the projections into the 2nd PC determine a pattern in terms of proximity regarding the months, i.e., projections of points with the same tag, e.g., J1, \dots , J8, always falls close to the same coordinate. This pattern repeats for all months and indicates, then, a periodicity of twelve months. Therefore, the first SSA-HJ-biplot combines different structural components of the TS CO₂, since the 1st PC is related to trend and the 2nd PC to seasonality.

Still in Figure 3, for any year, points near to each other indicate similarity in the behavior of the κ -lagged vectors, e.g., the set of points {A,Y,U} or {O,N,D}. It means that the κ -lagged vectors starting in April, May, and June, or the κ -lagged vectors beginning in October, November, and December resemble each other in terms of the object of interest. Also, the labeling strategy proved to be useful to capture the series behavior in the month itself, since April, May, and June correspond precisely to the periods in which the highest concentration of carbon dioxide occurs in the atmosphere. Besides, October, November, and December are the months with the lowest measured level of CO₂.

In turn, the column markers (ℓ -lagged vectors) are represented as black arrows up to the sixth ℓ -lagged vector (tagged as L1, \dots , L6 in Figure 3), ordered from top to bottom. From the seventh ℓ -lagged vector onwards, the pattern repeats itself, and so they were plotted in gray. It means that the first group of arrows, which is at the top, refer to the ℓ -lagged vectors beginning in January and July, just below those as starting in February and August, and so on. The angle between two consecutive arrows L_i and L_j , such that $i = 1, \dots, 5$ and $j = i + 1$, indicates a strong autocorrelation between the respective ℓ -lagged vectors since L_i and L_j form very sharp angles. Comparing the angles between L1 and the others up to L6, they vary from a value close to 0° to a value close to 90°, which suggests a fading of the autocorrelations. Figure 4 shows the SSA-HJ-biplot formed by the 2nd and 3rd PCs, while Figure 5 exhibits the SSA-HJ-biplot constructed from the 4th and 5th PCs. Along with the first SSA-HJ-biplot, these are

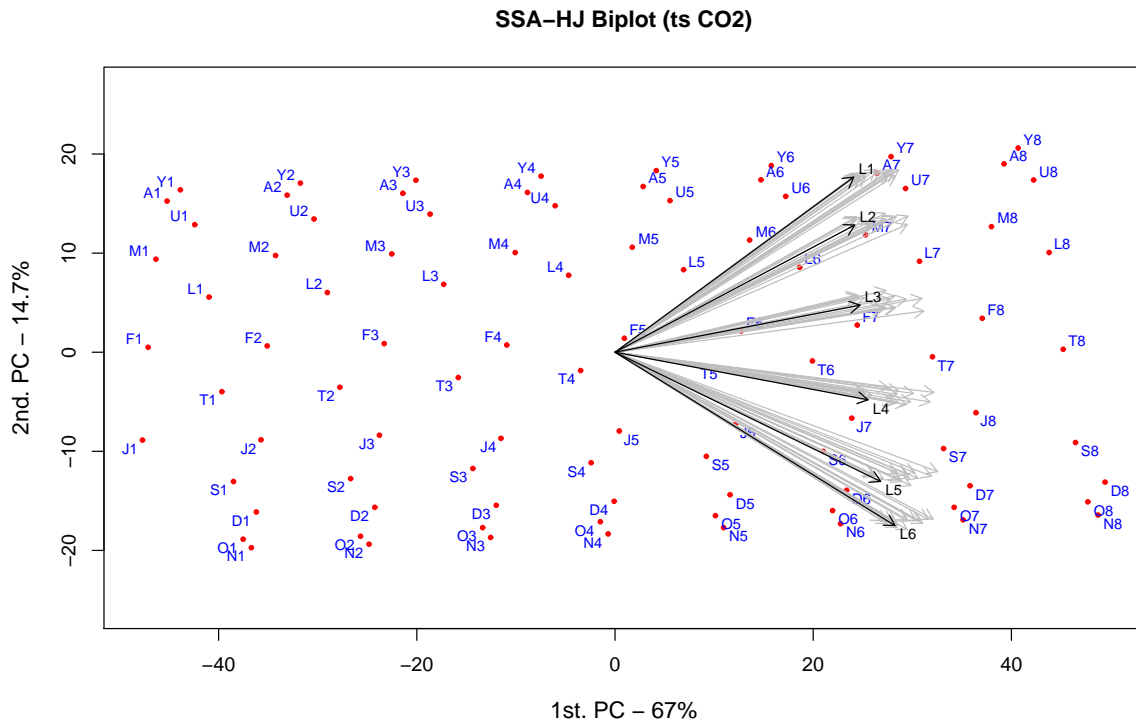


Figure 3. First SSA-HJ-biplot, constructed with the 1st and 2nd PC.

the only ones that produce interpretable results or evidence some pattern in the time series, being that these results are in agreement with the one verified in the scree plot of the singular values in Figure 2, where the pair of points related to $\sqrt{t'_2 t_2}$ and $\sqrt{t'_3 t_3}$ are around at the same level, the same concerning $\sqrt{t'_4 t_4}$ and $\sqrt{t'_5 t_5}$. In Figure 4, there are 12 distinct groups of row markers, each one of them referring to a κ -lagged vector starting in a specific month. Also, the column markers associated with each one of these groups show strong autocorrelation between the ℓ -lagged vectors. All of this indicates a seasonal pattern, with peaks and valleys separated by 12 months. In turn, the SSA-HJ-biplot in Figure 5 groups the lagged vectors two by two, e.g., January and July, February and August, and so on. Interpreting this together with the biplot in Figure 4, where these same groups occur but in the opposite directions, one can conclude that the valleys tend to be six months behind the peaks.

Therefore, the result of the grouping step for the decomposition of the TS CO2 should be \mathbf{X}_1 and \mathbf{X}_2 , the first corresponding to the trend component, and the second describing the seasonal component, in which

$$\mathbf{X}_1 = \sqrt{t'_1 t_1} t_1^* \mathbf{p}'_1,$$

and

$$\mathbf{X}_2 = \sum_{i=2}^5 \sqrt{t'_i t_i} t_i^* \mathbf{p}'_i,$$

with the remaining being related to the noise component. The application of the diagonal averaging procedure over \mathbf{X}_1 and \mathbf{X}_2 produces the reconstructed series $\hat{Y}^{(1)}$ and $\hat{Y}^{(2)}$, whose graphical representations are shown in Figure 6.

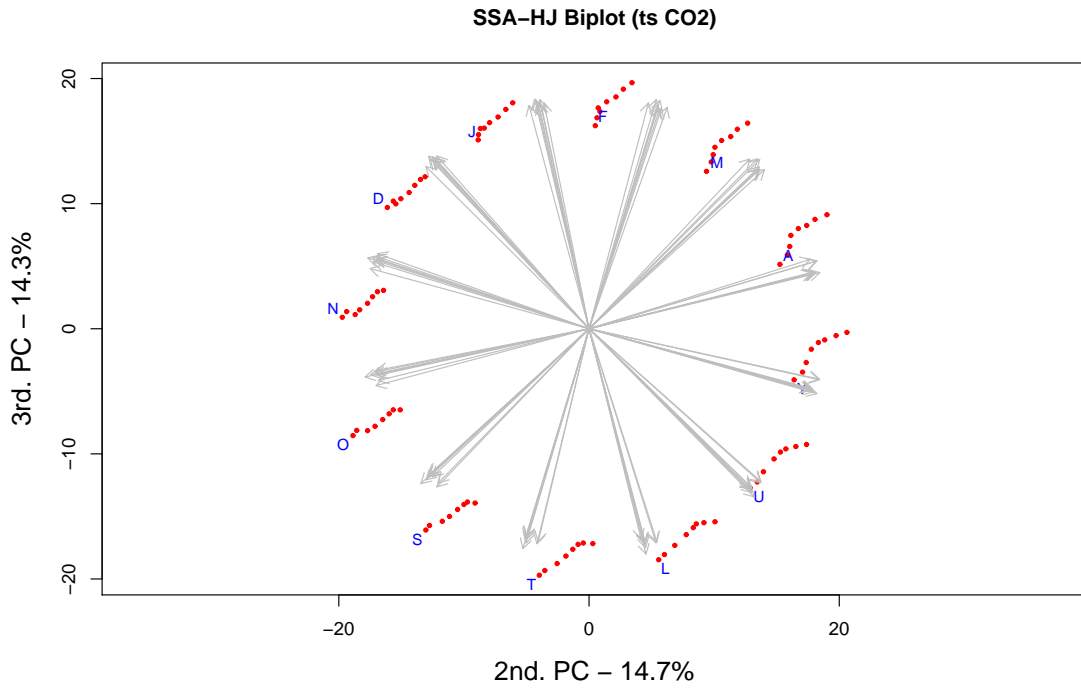


Figure 4. Second SSA-HJ-biplot of the 2nd and 3rd PCs describing an oscillatory component of TS CO2 of period 12

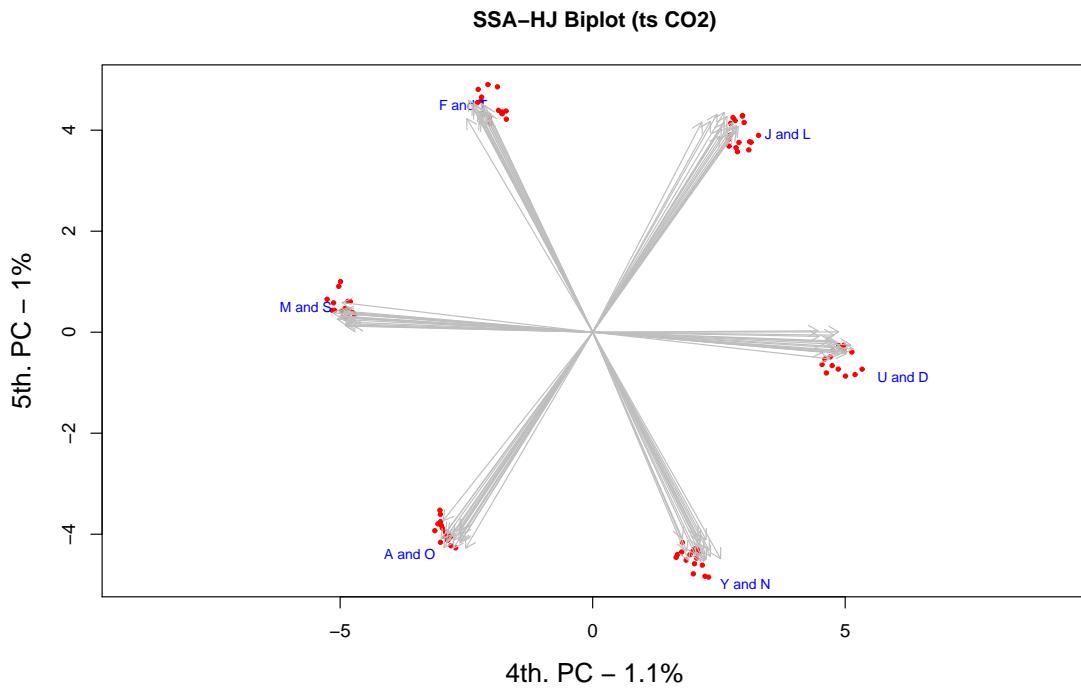


Figure 5. Third SSA-HJ-biplot of the 4th and 5th PCs also describing an oscillatory component of TS CO2.

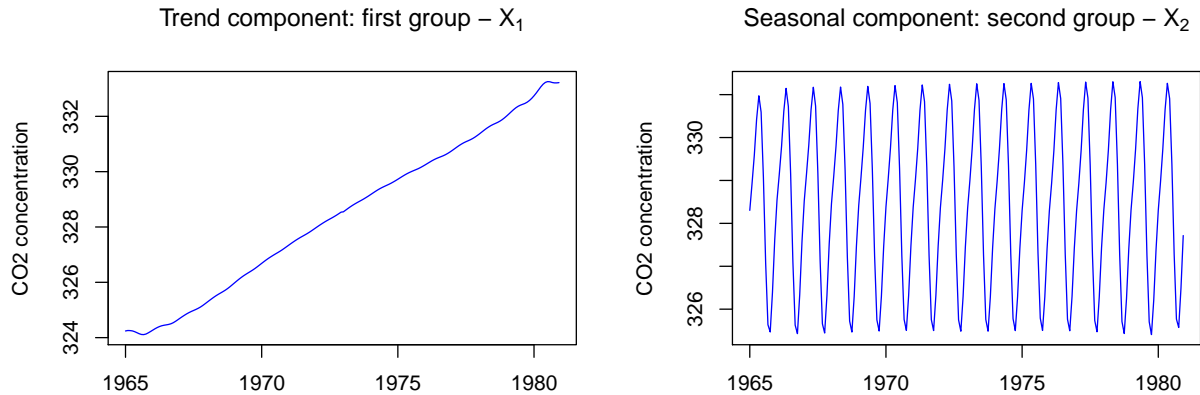


Figure 6. Separation of the trend and seasonal components of TS CO2 using the SSA-HJ-biplot technique.

5. Conclusions

This paper attempts to provide an integrative graphical tool to visualize and understand the underlying structure of the trajectory matrix, which is the result of the embedding step of the SSA. The SSA-HJ-biplot visualization method appears to be a promising exploratory technique, as it provides interpretability for the results of the SSA decomposition step, as illustrated by an example in this work. The SSA-HJ-biplots and auxiliary graphics provided a visual solution for the decomposition of the analyzed time series, properly separating the trend and the oscillatory component, using biplot axes up to the 5th PC. Also, it allowed the identification of all relevant eigentriple, composed by the singular values $\sqrt{t_i^* t_i}$, by the left singular vectors t_i^* , and by the right singular vectors p_i , $i = 1, \dots, 5$, to perform the grouping step. The study also revealed that the SSA-HJ-biplot points, representative of the row markers (j_i^*), can also depict the period itself in terms of dissimilarities, being possible to visually verify the months with the highest and lowest levels of CO2 concentration in the atmosphere throughout the years. The first SSA-HJ-biplot, built using the 1st and 2nd PCs, proved yet to be useful in dealing with autocorrelations between the column markers, which are drawn as arrows and represent the ℓ -lagged vectors. This study is promising in the sense that the SSA-HJ-biplot has great potential as an exploratory tool to analyze the structure of a univariate TS due to its visual appeal in such a complex issue. Nevertheless, TS may present complicated characteristics that make their analysis more challenging. For instance, prior detection of change-points in the TS is essential to highlight vital features, and consequently, to provide a better interpretation of SSA-HJ-biplots in complex TS data.

Acknowledgement

This work was supported by the Center for Research and Development in Mathematics and Applications (CIDMA) through the Portuguese Foundation for Science and Technology (FCT - Fundação para a Ciência e a Tecnologia), references UIDB/04106/2020 and UIDP/04106/2020.

REFERENCES

1. T. Alexandrov, *A method of trend extraction using Singular Spectrum Analysis*, REVSTAT, Statistical Journal, vol. 7, n. 1, pp. 1-22, 2009.
2. K. Gabriel, *The biplot graphic display of matrices with application to principal component analysis*, Biometrika, vol. 58, n. 3, pp. 453-467, 1971.
3. M.P. Galindo, *An alternative of simultaneous representation: HJ-biplot*, Questii, vol. 10, 1, pp. 13-23, 1986.
4. P. Geladi and B.R. Kowalsky, *Partial Least Squares regression: a tutorial*, Analytica Chimica Acta, vol. 185, pp. 1-17, 1986.
5. N. Golyandina, V. Korobeynikov and A. Zhigljavsky, *Singular Spectrum Analysis with R*, 1st ed., Springer, Berlin, 2018.
6. N. Golyandina, V. Nekrutkin and A. Zhigljavsky, *Analysis of Time Series Structure: SSA and Related Techniques*, 1st ed. Chapman & Hall/CRC, Boca Raton, Florida, 2001.
7. N. Golyandina and A. Shlemov, *Variations of Singular Spectrum Analysis for separability improvement: non-orthogonal decompositions of time series*, Statistics and its Interface, vol. 8, n. 3, pp. 277-294, 2015.
8. M. Greenacre, *Biplots in Practice*, FBBVA, Bilbao, Biscay, 2010.
9. A.B. Nieto, M.P. Galindo, V. Leiva and P.V. Galindo, *A methodology for biplots based on bootstrapping with R*, Colombian Journal of Statistics, vol. 37, n. 2, pp. 367-397, 2014.
10. S. Moritz, T. Bartz-Beielstein, *imputeTS: Time Series Missing Value Imputation in R*, The R Journal, vol. 9, n. 1, pp. 207-218, 2017. <https://doi.org/10.32614/RJ-2017-009>.
11. NOAA Homepage, <https://www.esrl.noaa.gov/gmd/ccgg/trends/>, last accessed 24/05/2019.
12. R Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2019 <https://www.R-project.org/>.
13. V.E. Vinzi and G. Russolillo, *Partial Least Squares algorithms and methods*, WIREs Comput Stat, vol. 5, pp. 1-19, 2013.
14. H. Wold, *Estimation of principal components and related models by iterative least squares*, in Multivariate Analysis, edited by P.R. Krishnaiah, Academic Press, New York, pp. 391-420, 1966.
15. S. Wold, C. Albano, W.J. Dunn, K. Esbensen, S. Hellberg, E. Johansson and M. Sjostrom, *Pattern recognition: finding and using regularities in multivariate data*, in Food Research and Data Analysis, edited by H. Martens and H. Russwurm, Applied Science Publishers, London, pp. 147-189, 1983.
16. K. Wright, *nipals: Principal Components Analysis using NIPALS or Weighted EMPCA, with Gram-Schmidt Orthogonalization*. R package version 0.7, 2020. <https://CRAN.R-project.org/package=nipals>