Ambiente & Água - An Interdisciplinary Journal of Applied Science
ISSN 1980-993X – doi:10.4136/1980-993X
www.ambi-agua.net
E-mail: ambi.agua@gmail.com

Use of singular spectral analysis to reconstruct average monthly flow time series with complex behavior

ARTICLES doi:10.4136/ambi-agua.2946

Received: 16 Jul. 2023; Accepted: 21 Nov. 2023

Michel Trarbach Bleidorn[1*]; Isamara Maria Schmidt[1]; José Antonio Tosta dos Reis[1]; Antônio Sérgio Ferreira Mendonça[1]; Wanderson de Paula Pinto[2]; Gutemberg Hespanha Brasil[3]

[1]Departamento de Engenharia Ambiental. Universidade Federal do Espírito Santo (UFES), Avenida Fernando Ferrari, n° 514, CEP: 29075-910, Vitória, ES, Brazil. E-mail: isamaraschmidt@gmail.com, jatreis@gmail.com, anserfm@terra.com.br
[2]Departamento de Ensino. Centro Universitário FAVENI (UniFAVENI), Rua do Rosário, n° 313, CEP: 07111-080, Guarulhos, SP, Brazil. E-mail: wandersondpp@gmail.com
[3]Departamento de Economia. Universidade Federal do Espírito Santo (UFES), Avenida Fernando Ferrari, n° 514, CEP: 29075-910, Vitória, ES, Brazil. E-mail: ghbrasil@terra.com.br
[*]Corresponding author. E-mail: michelbleidorn@gmail.com

ABSTRACT

Time series analysis is a useful tool for many practical water resource management applications, such as planning and anticipating conflict around water use. These analyses are also necessary for the use, design and operation of hydrotechnical works and for the protection of aquatic ecosystems. Basically, his review focuses on classic models such as Box and Jenkins. However, restrictive assumptions about data and residuals make its use difficult. As an alternative, non-parametric approaches are interesting due to their application versatility. One of the techniques that has been widely used in the analysis of flow time series is Singular Spectral Analysis (SSA). However, its application is concentrated as a pre-processor or in hybrid approaches and, because of this, many studies focus only on the model to be improved, leaving gaps in the process understanding that underlie the SSA. Within this context, and because of this, the objective of this study is to present the SSA technique in detail, through an application to a time series of monthly average flows with complex behavior (Jucu River, located in Southeast Brazil), in order to overcome gaps associated to a rigorous understanding of the procedures that make up the SSA. The Sequential SSA technique was used to decompose signal and noise components. The reconstructed series preserved the dynamics of the observed series, suggesting a strong influence of the signal component (trend and seasonality) on its behavior. It's expected that the technique can be used with greater fluidity by specialists and contribute to the management of water resources.

Keywords: decomposition, reconstruction, singular spectrum analysis.

Emprego da análise espectral singular para a reconstrução de séries temporais de vazões médias mensais com comportamento complexo

RESUMO

A análise de séries temporais é uma ferramenta útil para muitas aplicações práticas de gestão de recursos hídricos, como no planejamento e antecipação de conflitos em torno do uso d'água. Estas análises são também necessárias para a concepção e operação de obras hidrotécnicas e para a proteção de ecossistemas aquáticos. Basicamente, sua análise se concentra em modelos clássicos, como Box e Jenkins. No entanto, suposições restritivas sobre dados e resíduos dificultam seu uso. Como alternativa, abordagens não paramétricas são interessantes devido a sua versatilidade de aplicação. Uma das técnicas que têm sido amplamente utilizada na análise de séries temporais de vazão é a Análise Espectral Singular (SSA). No entanto, sua aplicação é concentrada como pré-processador ou em abordagens híbridas e por conta disso, muitos estudos se concentram apenas no modelo a ser aprimorado, deixando lacunas de entendimento dos processos que fundamentam a SSA. Dentro desse contexto, e por conta disso, o objetivo deste estudo é apresentar detalhadamente a técnica SSA, através de uma aplicação à uma série temporal de vazões médias mensais com comportamento complexo (rio Jucu, localizado no Sudeste do Brasil), a fim de superar lacunas associadas ao entendimento rigoroso dos procedimentos que compõem a SSA. A SSA Sequencial foi utilizada para decompor os componentes sinal e ruído. A série reconstruída preservou a dinâmica da série observada, sugerindo forte influência do sinal (tendência e sazonalidade) em seu comportamento. Espera-se que a técnica possa ser usada com maior fluidez pelos especialistas e contribuir na gestão dos recursos hídricos.

Palavras-chave: análise espectral singular, decomposição, reconstrução.

1. INTRODUCTION

Knowledge of river-flow behavior is fundamental to the understanding of the hydrological regime that a watershed presents as a result of changes in land use and occupation and climatic conditions. This information is important for water resource planning. In this context, flow time series analysis tools are used to produce information that can assist in the elaboration of water resource plans, a relevant instrument established by the Brazilian National Water Resources Policy, instituted by Law No. 9,433 (Brasil, 1997).

A time series is defined as a sequence of observations of any random phenomenon ordered in time (Reisen *et al.,* 2008; Pinto *et al.,* 2018). In river flows, it is composed of stochastic and deterministic (or harmonic) components (Salas *et al.,* 1980; Apaydin *et al.,* 2021). The stochastic component represents a sequence of observations and random relationships defined by probability functions that change with time. The deterministic component, on the other hand, is composed of trend and seasonality components (Meng *et al.,* 2019), and can be reproduced by mathematical functions that are based on past behavior, presenting a known pattern. Basically, a time series can be analyzed by two approaches: parametric and non-parametric.

In the first approach, called “classical analysis”, the time series is analyzed in terms of its trend, seasonality and noise components, in order to understand the series-generating mechanism, extract relevant periodicity in the observations, describe its behavior and make predictions (Bayer and Souza, 2010). Its application was widespread after statisticians Box and Jenkins developed the methodology that became known as “ARMA” (Autoregressive and Moving Averages) class time series modeling in the 1970s. Since then, it has become a tool widely used for forecasting river flow (Abudu *et al.*, 2010). Some examples in Brazil include the studies by Bayer *et al.* (2012) who fitted the SARIMA $(3,0,0)(2,1,2)_{12}$ model to data on average monthly flows from the Potiribu River, a tributary of the Ijuí River, in Rio Grande do Sul, Brazil; Pinto *et al.* (2015) who, through a comparative perspective, adjusted the SARIMA $(1,1,1)(1,1,2)_{12}$ model, which proved to be the most appropriate for forecasts of average monthly flows of the Doce River in Colatina, Espírito Santo, Brazil ; and Bleidorn *et al.* (2019) who verified the best fit of the SARIMA model $(1,0,0)(5,1,0)_{12}$ for forecasts of average monthly flows of the Jucu River in Domingos Martins, Espírito Santo, Brazil.

One of the biggest difficulties of time-series analysis in the time domain focuses on how noise is interpreted, since it is considered an important component in the analysis. In addition, the approach makes some distributive assumptions of the data, such as normality, also with respect to residuals, in addition to no autocorrelation. However, these assumptions are difficult to achieve due to the non-linear river's complexity, consequence of the biotic and abiotic interrelationships of watersheds (Xie *et al.*, 2016; Du *et al.*, 2017; Unnikrishnan and Jothiprakash, 2018; Meng *et al.*, 2019; Apaydin *et al.*, 2021).

The second approach presents interesting requirements in that it doesn't impose distributional restrictions on the data and its analysis isn't based on construction of *a priori* model. Among the available methodologies the Singular Spectrum Analysis (SSA) is considered a powerful tool for time series analysis because it incorporates elements of multivariate statistics, classical time series analysis, multivariate geometry, dynamical systems and signal processing (Golyandina and Zhigljavsky, 2013). The main objective of the SSA technique is to decompose any original series with an apparently complex pattern into a sum of signals (trend and harmonics) and noise (Hassani, 2007; Hassani and Zhigljavsky, 2009, Hassani *et al.,* 2013). Unlike classical analysis, in SSA, noise is removed for analysis and, due to this procedure, it has been used as a tool for pre-processing hydrological data, with the main objective, which was very successful, of improving the performance of different models for predictions (Unnikrishnan and Jothiprakash, 2018).

As an example, Wu *et al.* (2010) verified that the application of the modular Artificial Neural Network (ANN) has its accuracy increased after the application of SSA in precipitation series from watersheds located in China and India. Unnikrishnan and Jothiprakash (2018) found that SSA improved the performance of ANN models both in predicting single and multiple time steps ahead of daily rainfall time series for Koyna watershed, in India. (Wang *et al.*, 2019a) observed that dual processing, through SSA and Adaptive Noise Enhanced Full Set Empirical Mode Decomposition (ICEEMDAN), allowed an even more significant improvement in the performance of the extreme learning machine model, in an application to the monthly flow data from Gulang River basin, China. Apaydin *et al.* (2021) observed that the SSA decomposition increased the monthly accuracy of flow forecast by ANN, Convolutional Neural Network (CNN) and Long-Short Term Memory networks (LSTM), considering data from the Sakarya basin, Turkey, by 24.11%, 18.40% and 5.11%, respectively.

An alternative that has also been showing good performance is hybrid modeling. Unnikrishnan and Jothiprakash (2020) combined SSA, ARIMA (Autoregressive Integrated Moving Averages) and ANN in a case study that considered daily rainfall data from the Koyna watershed, India. Zubaidi *et al.* (2018) coupled SSA with machine learning (LS-SVR) and random forest (RF) for monthly rainfall forecast in two watershed reservoirs (Deji and Shihmen) located in Taiwan. Pham *et al.* (2021) found that coupling SSA with a single least square support vector machine (LS-SVM) significantly increased the accuracy of drought forecasts for the Standardized Precipitation Index (SPI), considering data obtained in Taiwan. Apaydin *et al.* (2021) found that the coupling of SSA and ANN increased prediction accuracy efficiency for a river located in Taiwan by 24.11%.

As observed, current literature has focused efforts on applying the SSA technique through hybrid modeling or as a preprocessor. In both cases, its application has been useful to overcome difficulties associated with the noise component and increasing efficiency gains of forecasts. Because of this, many studies focus only on the model to be improved, leaving gaps in understanding the processes that underlie SSA. Furthermore, the definition of only two attributes (window length and separability), the ability to decompose series into interest components, the noise removal and the non-establishment of restrictive assumptions on the data, give potential to the aforementioned technique. In this context, this study presents the SSA technique in detail, through an application to a monthly average flow time series with complex behavior (Jucu River, located in Southeastern Brazil), in order to overcome gaps associated with rigorous understanding of the procedures that make up SSA. It is expected that the results presented in this research can encourage the use of SSA in flow time series, helping specialists and managers to understand water body regimes, offering additional support to water-resource management.

2. MATERIALS AND METHODS

2.1. Study Area

The Jucu River Watershed (Figure 1) is located in the Center-South region of Espírito Santo state, Brazil. The Jucu River is 166 km long and has a drainage area of 2.032 km² (Reisen *et al.*, 2008). The Jucu River begins at the junction of the Jucu Braço Sul and Jucu Braço Norte Rivers and flows into Barra do Jucu Beach, in the municipality of Vila Velha.

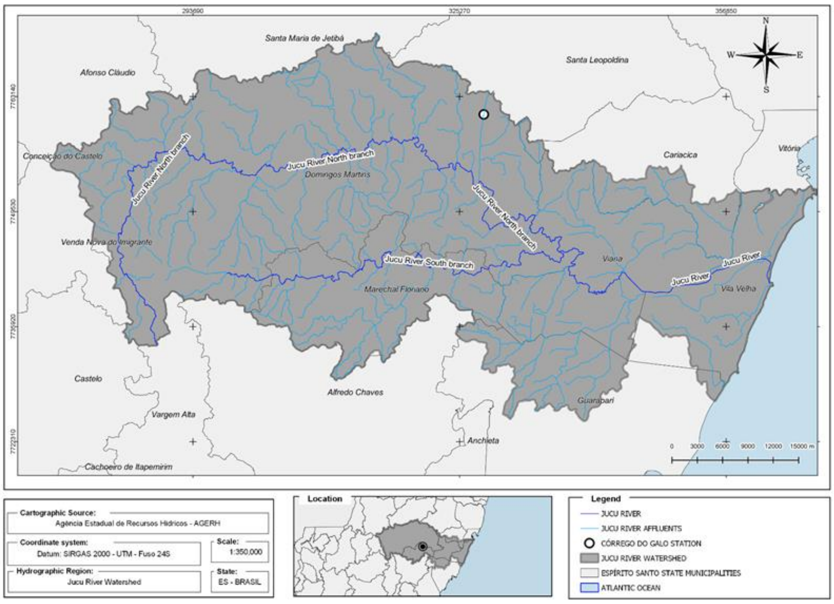

Figure 1. Jucu River Watershed.

2.2. Data

The average monthly flow time series was obtained from the Hydrological Information System (HIDROWEB) maintained by the National Water Agency (ANA, 2022). The Jucu River monitoring station whose data was considered is called "Córrego do Galo" (Braço Norte), is located in the municipality of Domingos Martins, Espírito Santo, Brazil and is shown in Figure 1. This station drainage area corresponds to 980 km² and has been in operation since 1969. In this study, consistent data up to the year 2014 were considered, resulting in a 46-year-long historical series.

2.3. Singular Spectrum Analysis (SSA)

SSA is a powerful non-parametric method for time series analysis that can efficiently identify and extract trends, seasonal behavior and noise (Zhang *et al.*, 2017). The standard SSA consists of two main steps: (1) decomposition and (2) reconstruction. Each step in turn presents two phases: (1.1) embedding and (1.2) singular value decomposition (SVD); and (2.1) grouping and (2.2) diagonal averaging.

$Y_N = (y_1, y_2 \dots, y_N)$ be a one-dimensional series. The mathematical theory of the SSA technique can be described as presented below, based on Golyandina *et al.* (2001) and Hassani (2007).

Embedding: Transfers the one-dimensional series $\mathbb{Y}_N$ into a multidimensional series $X_1, \dots, X_K$, with vectors $X_I = (y_I, \dots, y_{i+L-1})^T \in \mathbb{R}^L$, where $K = N - L + 1$. Vectors X_I are called *lagged vectors*. The single parameter of the embedding step is the window length L, an integer such that $2 \leq L \leq N$. The result of this step is the trajectory matrix X shown in Equation 1 below. The trajectory matrix has the property of a Hankel matrix, since the $(i,j)_{th}$ element of the matrix X is $x_{ij} = y_{i+j-1}$, resulting in equal elements in the antidiagonals $i + j = const$.

$$\mathbf{X} = [X_1, \dots, X_K] = (x_{ij})_{I,j=1}^{L,K} = \begin{bmatrix} y_1 & y_2 & y_3 & \cdots & y_K \\ y_2 & y_3 & y_4 & \cdots & y_{K+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_L & y_{L+1} & y_{L+2} & \cdots & y_N \end{bmatrix} \quad (1)$$

Singular Value Decomposition (SVD): after embedding, SVD is applied to matrix X and as a result the matrix is converted into several small matrices called Eigentriples which contain a certain energy of the series (signal). The matrix S is the result of multiplying the trajectory matrix X by its transpose, X^T, that is, $S = XX^T$. Applying SVD to S is obtain the eigenvalues $\lambda_1, \dots, \lambda_L$ arranged in decreasing order according to their magnitudes $(\lambda_1 \geq \cdots \lambda_L \geq 0)$ and the corresponding eigenvectors $U_1, \dots, U_L$. The result of the singular value decomposition of the trajectory matrix X can be written as Equation 2:

$$\mathbf{X} = \mathbf{X_1} + \dots + \mathbf{X_d} \quad (2)$$

Where, d is the number of non-zero eigenvalues of the matrix **S**, $\mathbf{X}_I = \sqrt{\lambda_i} U_i V_i^T$ $(i = 1, \dots, d)$ and $V_1, \dots, V_d$ are principal components defined as $V_i = \frac{\mathbf{X}^T U_i}{\sqrt{\lambda_i}}$. The $\mathbf{X}_I$ matrix is of unitary rank (rank-one), often referred to as elementary matrix. The collection $\sqrt{\lambda_i} U_i V_I$ is called the $i - th$ eigentriple of the trajectory matrix **X** and λ_i is its singular value, or spectrum.

Grouping: the objective of this step is to select significant Eigentriples. In this step, the elementary matrix X_I $(i = 1, \dots, d)$, are grouped into m subsets of $I_1, \dots, I_m$. Setting $I = \{i_1, \dots, i_p\}$, then the resulting matrix X_I corresponding to group I is defined as $X_I = X_{i1} + \dots + X_{ip}$. These matrices are calculated for $I = I_1 + \cdots + I_m$ and the expansion shown in Equation 2 leads to the decomposition shown in Equation 3:

$$\mathbf{X} = \mathbf{X}_{I1} + \dots + \mathbf{X}_{Im} = \sum_{i \in I} \mathbf{X}_i + \varepsilon \quad (3)$$

The procedure for choosing the set of $I = I_1, \dots, I_m$ is called grouping the Eigentriples. The choice of sets $I_1, \dots, I_m$ is the second and last decision necessary for the application of the SSA method, which is based on the property called separability, using some method, such as principal components analysis, single vector graphics or hierarchical grouping. In this work, the weighted correlation was considered, calculated as follows: the weighted correlation (or commonly called *w-correlation*) measures the dependence between two subseries $\mathbb{Y}_N^{(i)}$ and $\mathbb{Y}_N^{(j)}$ and can be expressed by Equation 4.

$$\rho^{(w)}\left(\mathbb{Y}_N^{(i)}, \mathbb{Y}_N^{(j)}\right) = \frac{\left(\mathbb{Y}_N^{(i)}, \mathbb{Y}_N^{(j)}\right)_w}{\left(\sqrt{\left(\mathbb{Y}_N^{(i)}, \mathbb{Y}_N^{(i)}\right)_w} \sqrt{\left(\mathbb{Y}_N^{(j)}, \mathbb{Y}_N^{(j)}\right)_w}\right)} \quad (4)$$

Where, $\left(\mathbb{Y}_N^{(i)}, \mathbb{Y}_N^{(j)}\right)_w = \sum_{p=1}^{N} w_p y_p^{(i)} y_p^{(j)}$, $(I, j = 1,2)$. Hassani and Mahmoudvand

(2018) indicate that if the absolute value of the weighted correlation (w) is small, the corresponding series are almost orthogonal; however, if it is large, the two series are far from orthogonal and are called weakly separable. It is important to remember that the grouping of $\mathbf{X}_I$ is related to the signal of the series, which contains relevant information such as trends and oscillations, and ε indicates the error term of the Eigentriples.

Diagonal averaging: after grouping, the grouped matrices (Eigentriples) are transformed into a time series using diagonal averaging. In this procedure, the matrix elements are calculated over the antidiagonals to obtain the time series. Thus, at the end of the diagonal averaging, a new time series is generated. Let Y be a matrix $L \times K$ with elements y_{ij}, $1 \leq i \leq L$, $1 \leq j \leq K$. Defining $L^* = min(L,K)$, $K^* = max(L,K)$ and $N = L + K - 1$. Let $y^*_{ij} = y_{ij}$ if $L < K$, $y^*_{ij} = y_{ji}$ if $L > K$ and, taking the diagonal average, the matrix Y is transferred into the series $(\tilde{y}_1, \tilde{y}_2 \ldots, \tilde{y}_N)$, using Equation (5):

$$y_k = \begin{cases} \frac{1}{k}\sum_{m=1}^{k} y^*_{m,k-m+1} & \text{para } 1 \leq k < L^*, \\ \frac{1}{L^*}\sum_{m=1}^{L^*} y^*_{m,k-m+1} & \text{para } L^* \leq k \leq K^*, \\ \frac{1}{N-k+1}\sum_{m=k-K^*+1}^{N-K^*+1} y^*_{m,k-m+1} & \text{para } K^* < k \leq N. \end{cases} \quad (5)$$

This corresponds to averaging the matrix elements over the "antidiagonals". Diagonal averaging (5) applied to a resultant matrix $\mathbf{X}_{I_k}$ produces a reconstructed series like as $\tilde{\mathbb{Y}}^{(k)} = \left(\tilde{y}_1^{(k)}, \ldots, \tilde{y}_N^{(k)}\right)$. Therefore, the initial series $\mathbb{Y}_N = (y_1, y_2 \ldots, y_N)$ is decomposed into a sum of m reconstructed series (Golyandina and Korobeynikov 2014), using Equation 6:

$$y_n = \sum_{k=1}^{m} \tilde{y}_n^{(k)}, (n = 1, 2, \ldots, N) \quad (6)$$

2.4. Performance Indexes

To evaluate the performance of the reconstruction of the series, the performance indicators BIAS, Root Mean Square Error (RMSE), Mean Absolute Error (MAE), agreement index (d_2) and Pearson's correlation coefficient (ρ), represented by Equations 7, 8, 9, 10 and 11, respectively, were used.

BIAS: Quantifies under and overestimation in relation to average observations.

$$BIAS = \frac{1}{n}\sum_{i=1}^{N}(x_i - \tilde{x}_i) \quad (7)$$

RMSE: Root Mean Square Error is defined as the mean square difference between predicted or adjusted values and the respective true values. In general, lower values indicate good performance.

$$RMSE = \frac{1}{n}\sqrt{\sum_{i=1}^{N}(x_i - \tilde{x}_i)^2} \quad (8)$$

MAE: The Mean Absolute Error is a measure of the accuracy of the fit or prediction of a model. Smaller values indicate good performance.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{x_i - \tilde{x}_i}{x_i}\right| \quad (9)$$

d_2: The index of agreement reflects the agreement between the fit of the model and the observed data, with a desirable value of 1, which indicates perfect agreement.

$$d_2 = 1 - \left[\frac{\sum_{i=1}^{n}(x_i - \tilde{x}_i)^2}{\sum_{i=1}^{n}(|x_i - \bar{\tilde{x}}| + |\tilde{x}_i - \bar{\tilde{x}}|)}\right] \quad (10)$$

ρ: The Pearson correlation coefficient describes the relationship between the variables and those closest to the extremes (-1 or 1) represent strong correlations.

$$\rho = \frac{\sum_{i=1}^{n}[(x_i - \bar{x})(\tilde{x}_i - \bar{\tilde{x}})]}{\sum_{i=1}^{n}\hat{\sigma}(x_i)\,\hat{\sigma}(\tilde{x})} \quad (11)$$

2.5. Computational resources

Electronic spreadsheets were used to organize the data and the R software (R Development Core Team, 2018) to carry out the analysis of the SSA implemented in the RSSA package. The RSSA package presents an efficient implementation of the main SSA algorithms, containing many visual tools that are useful to make the proper choice of SSA parameters and examine the results (Golyandina and Korobeynikov, 2014).

3. RESULTS AND DISCUSSION

3.1. Descriptive statistics

For an initial understanding, Table 1 presents some descriptive measures of the average monthly flow variable. The mean of the variable corresponds to 13.86 $m^3.s^{-1}$ with standard deviation 7.47 $m^3.s^{-1}$ and coefficient of variation 53.89%. The high values of the standard deviation and coefficient of variation indicate that the average has little representation, a characteristic that can be associated with the presence of seasonality in the series. Values of 1.89 for skewness and 4.90 for kurtosis indicate that the distribution does not follow a Normal Distribution and has heavier tails than this one.

Table 1. Descriptive measures of Jucu River average monthly flow variable.

Descriptive measures	E1
Average ($m^3.s^{-1}$)	13.86
Median ($m^3.s^{-1}$)	11.62
Coefficient of Variance (%)	53.89
Minimum Value ($m^3.s^{-1}$)	5.13
Maximum Value ($m^3.s^{-1}$)	56.50
Kurtosis	4.90
Asymmetry	1.89

Figure 2 indicates the temporal evolution of the series. As indicated by the descriptive statistics, the seasonal characteristic is noticeable due to the well-defined intra-annual variability, with periods of flood and drought.

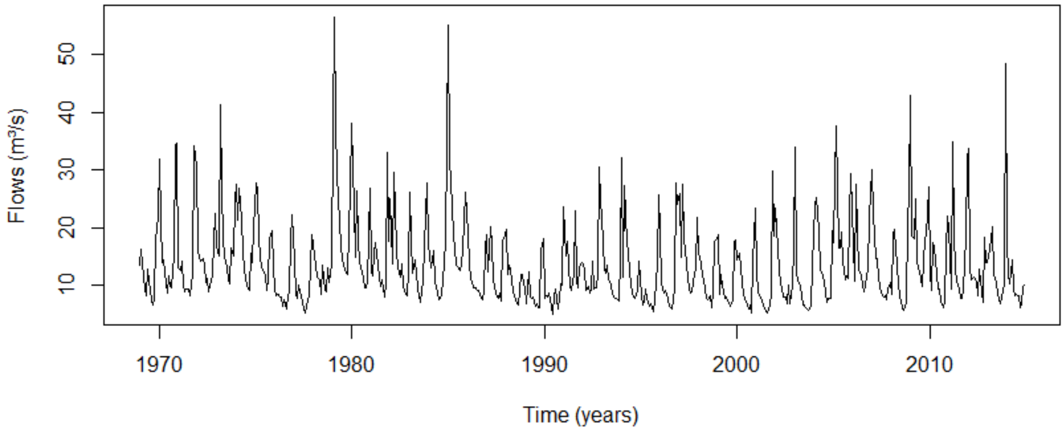

Figure 2. Jucu River time series average monthly flows.

3.2. Decomposition of the original series

Ballot (1847) made the first attempt to decompose time series into trend and seasonality back in 1847, modelling the trend by a polynomial and the seasonality by dummy variables. Macaulay (1931) was concerned with extracting seasonality through the smoothing of time series. Petropoulos and Apiletti (2022) reports that the main idea of this method comes from the observation that averaging a time series with a window size equal to the seasonal period leaves the trend almost intact, while effectively removing seasonal and noisy components.

The main purpose of SSA is to decompose the observed time series into a sum of time series so that each can be identified as a trend, periodicity, quasi-periodicity or noise component (Abdollahzade *et al.*, 2015). Equation 12 presents the components of a time series (S_t).

$$S_t = S_T + S_S + S_C + \varepsilon_t \tag{12}$$

Where, S_t represents the trend, S_S represents the seasonality, S_C represents the cycle and ε_t the noise (stochastic component).

Based on Morettin and Toloi (2006) and Komorník *et al.* (2006) the descriptions of the components are presented below:

- Trend: long-term component that represents growth, decline or stability in a time series by which the speed with which these changes occur is identified.

- Seasonal component: a cyclical pattern of change in data that repeats from year to year (intra-annual). Seasonal phenomena are associated, for example, with the seasons.

- Cyclical component: wave fluctuation around the trend (inter-annual). El Niño and La Niña phenomena typically have higher occurrences.

- Noise component: the residual of the values after the removal of the other components, characterized by ascending or descending movements, of great instability and randomness.

The series decomposition step is defined using the window length (L) and, consequently, the lag operator (K). According to the authors Golyandina *et al.* (2001), Hassani (2007) and Golyandina and Korobeynikov (2014), L must be a number divisible by the period and close to half the size of the series ($L \cong \frac{N}{2}$). Based on descriptive statistics and in the study by Bleidorn *et al.* (2019), Jucu River has a 12-month seasonality. Because of this, the window size was initially fixed at L=276, resulting in a lag operator K=277.

3.3. Separation of signal and noise

As demonstrated, the SSA technique allows decomposing a time series into signal and noise. The term commonly called "signal" represents the deterministic components of interest, such as trend and sine wave (Hassani, 2007). The grouping of the series components is done with the aid of the graphs of the singular values of the eigenvectors (scatter plot) and weighted correlation (w-correlation), presented in Figures 3 and 4, respectively.

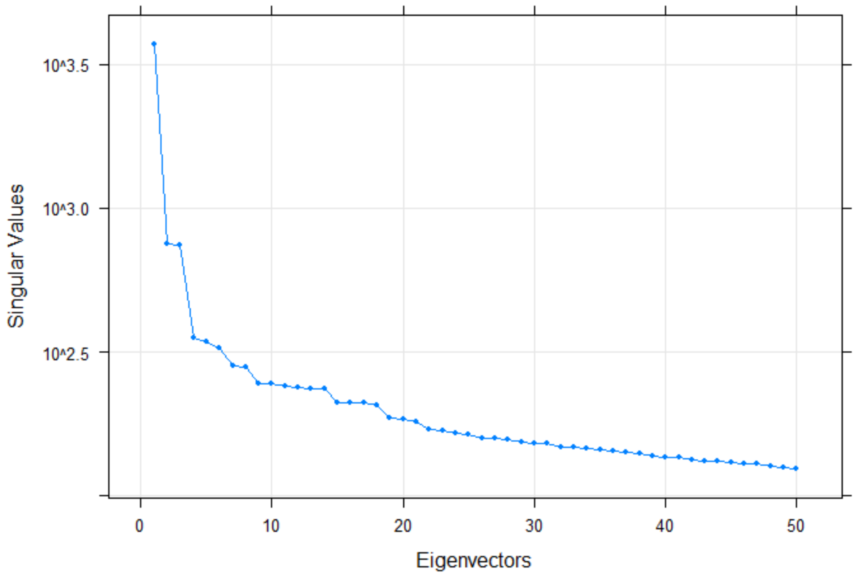

Figure 3. Singular values resulting from the decomposition of the Jucu River original series (L=276).

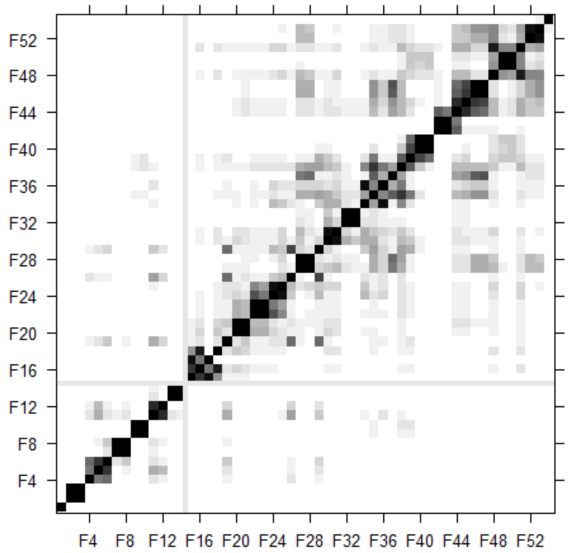

Figure 4. Weighted correlation matrix of the first 52 autovectors.

Golyandina and Korobeynikov (2014) indicate how the scatterplot can be useful in extracting the signal and noise components:

- Isolated vectors: presents only one point (1D) isolated from eigenvectors U_i that indicate trend;

- Paired: also called "plateau", is represented by 2D graphics, comprising eigenvectors

U_{i+1} that signal sine waves (period); and,

- Smooth: vectors with smooth decay, indicating the noisy part of the series.

According to Golyandina *et al.* (2018), w-correlation contains very useful information that can be used for separability detection and group identification. Well separated components have weak (or zero) correlation, while poorly separated components usually have high correlation with other eigenvectors.

As shown in Figures 3 and 4, eigenvectors 1, 4, 5 and 6 represent the trend component. The sine wave component can be formed by components 2-3, 7-8, 11-12 and 13-14, for forming plateaus in the scatter plot and by the strong correlation between them in the w-correlation plot. Clearly can be observed the noise presence from eigenvector 15, associated mainly with its smooth decay observed in Figure 4 and the presence of eigenvectors correlations with other eigenvectors, depicted in Figure 5. However, it is also possible to observe that the signal components also correlate with other signal and noise components in the series, giving rise to the problem of "weak separability".

According to Salas *et al.* (1980), this is due to the fact that time series, including hydrological ones, are considered to be composed of stochastic and periodic terms superimposed on general trends. In a complementary way, as the seasonal components are approximately orthogonal to the slowly varying components, one can consider the existence of a mixing problem due to the lack of strong separability (Golyandina and Korobeynikov, 2014). This consideration is pertinent when the series studied presents a complex behavior, as can be observed in the trend line (Figure 5). This condition is due to the existence of accentuated non-linear trends and distinct change points in the direction of the trend throughout the recording period (Aswathaiah e Nandagiri, 2020).

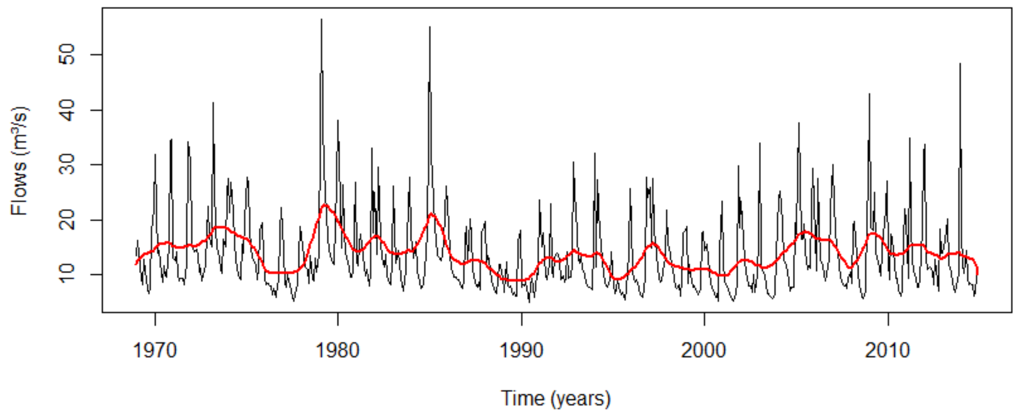

Figure 5. Jucu River observed series and trend lines.

The complexity of the series studied makes it difficult to fully decompose it. As an alternative, Golyandina and Korobeynikov (2014) recommend the technique called Sequential SSA, which differs from the basic SSA in the decomposition step, by conforming two stages (sequential): (i) the trend is extracted with the smallest possible window length divisible by the period. In this study, $L = 12$ was adopted because it was the smallest window length value that met the aforementioned criterion, a choice also made by Du *et al.* (2017); and, (ii) later, the periodic components are detected and extracted from the residue with ($L \cong \frac{N}{2} = 276$). Applying this new approach, a new scatter plot is formed (Figure 6). Based on the presentation and discussion in Figure 4, the isolated eigenvector extracted by the first stage of Sequential SSA represents the trend component. The second stage of Sequential SSA allows the extraction of sinusoidal components, formed by the plateaus depicted in Figure 6.

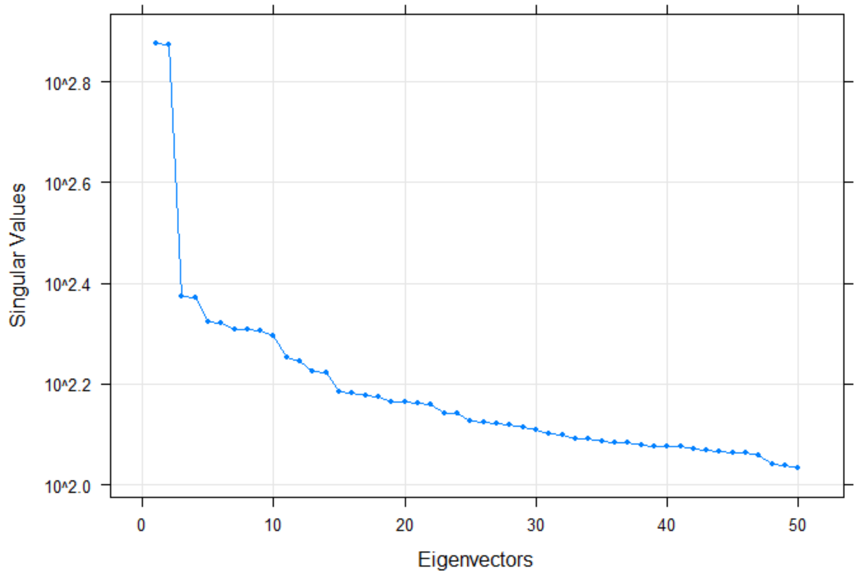

Figure 6. Singular values resulting from the decomposition of the Jucu River series (L= 276) without trend – harmonic components.

Applying Sequential SSA, a new conformation is also observed in the w-correlation graph (Figure 7). The first components represent sine waves, as they form pairs of eigenvectors that correlate with each other. Furthermore, Figure 7 also demonstrates that sine wave components are not correlated with other eigenvectors. Therefore, it is inferred that the application of Sequential SSA is an appropriate decomposition alternative when dealing with the complex behavior observed in the studied series. As discussed, the first extracted component comprises the trend component. Based on the analyses of the scatterplots and w-correlation (see Figures 6 and 7), the eigenvector pairs 1-2 (F1-F2), 2-3 (F3-F4) and 5-6 (F5-F6) represent sine wave components. Due to the relatively smooth decay of singular values after eigenvector 7 (Figure 6) and the presence of correlations between them (Figure 7), the eigenvectors from 7 to 276 represent the series noise.

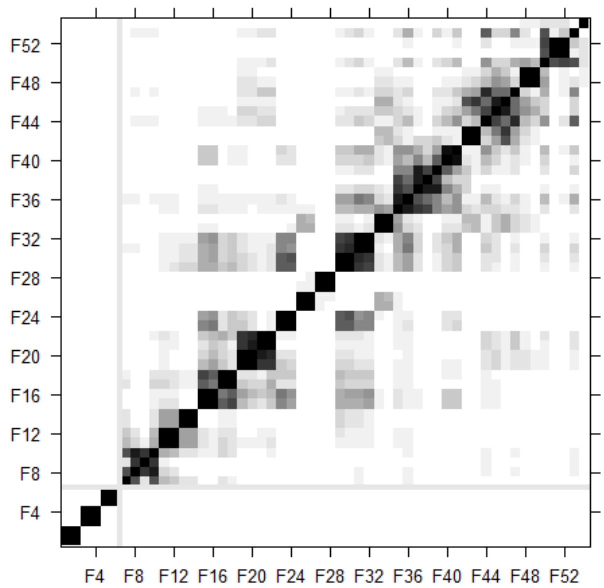

Figure 7. First 52 autovectors weighted correlation matrix after trend extraction.

Figure 8 presents the eigenvector pairs. These graphs are important for illustrating harmonic components (seasonality) through the identification of geometric shapes. In the eigenvector pairs 1-2, there is a well-defined dodecagon. The seasonal period is determined by the number of edges that make up the geometric shape. Therefore, this eigenvector pair presents a seasonality with a 12-month period. The eigenvector pair 3-4 presents a seasonal period equal to 6 months. Further, the 5-6 eigenvector pair represents another 4-month seasonal component. These conformations reproduce a characteristic of strong intra-annual variability. The first (12 months) may involve repeating monthly behavior, year after year; the second (6 months) can represent the variability of drought and flood; and the last (4 months), are possibly influenced by the seasons.

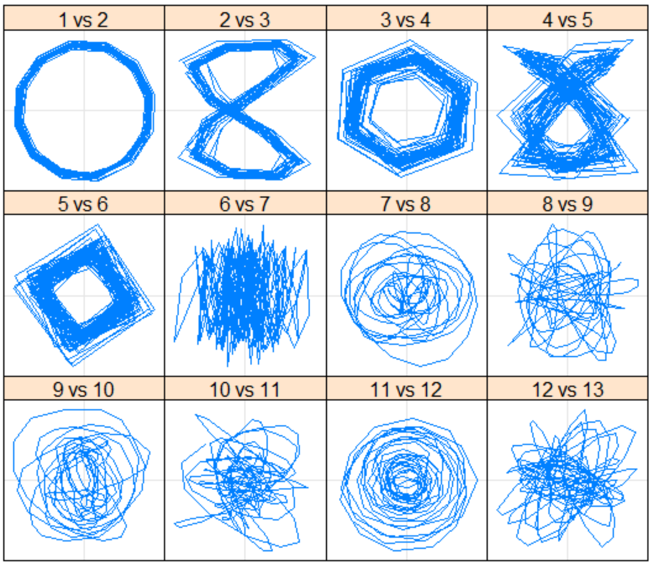

Figure 8. Jucu River monthly flows series eigenvectors pairs.

3.4. Series reconstruction

In this step, the series is reconstructed considering the components extracted in the decomposition step. Eigenvectors 1 for trend and 1 to 6 for seasonality were considered. Figures 9 and 10 demonstrate the reconstructed series considering the trend and seasonality components, respectively. It is possible to notice that the residues extracted by Step 1 (trend decomposition) present seasonal components. In Figure 10, it is observed that after trend and seasonality decomposition, the residues present random distribution, representing the series noisy part. In watersheds, the noise present in a flow series can be associated with land use and occupation conditions and local climate changes (Ren *et al.*, 2018; Malekian *et al.*, 2019).

The extreme values are largely observed in the trend component decomposition, which is consistent with the result obtained by Apaydin *et al.* (2021). Aswathaiah and Nandagiri (2020) also verified good performance of the SSA technique in extracting the trend component in precipitation data from a watershed in India. The seasonality component was not affected by extreme events, which shows that the decomposition technique adequately portrayed seasonality within the series. This characteristic indicates that the technique may contribute with risk-management plan definitions, environmental preservation of rivers and extreme event mitigation.

In addition, flow fluctuations influence the temporal distribution of runoff and sediment transport throughout the year which, in turn, plays a significant role in the social economy and ecology (Ran *et al.*, 2020). Ren *et al.* (2018) also point out that knowing the variability of the river, richer hydrological information can be obtained for ecological conservation and conflicts between the water demand of ecosystems and humans can be mitigated, especially in dry years.

Pruski *et al.* (2014), Silva *et al.* (2015) and Bleidorn *et al.* (2023) portray how seasonality is an important factor in environmental and economic management in the concession of water grants in a sub-basin of the Paracatu River and in the watersheds of the Paraobepa and Doce Rivers, respectively.

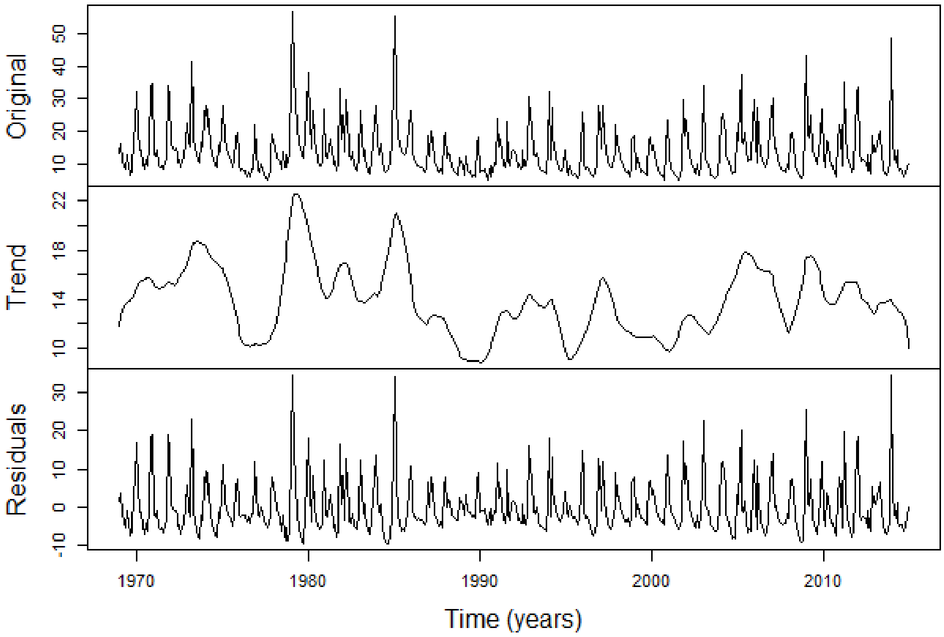

Figure 9. Jucu River series trend reconstruction.

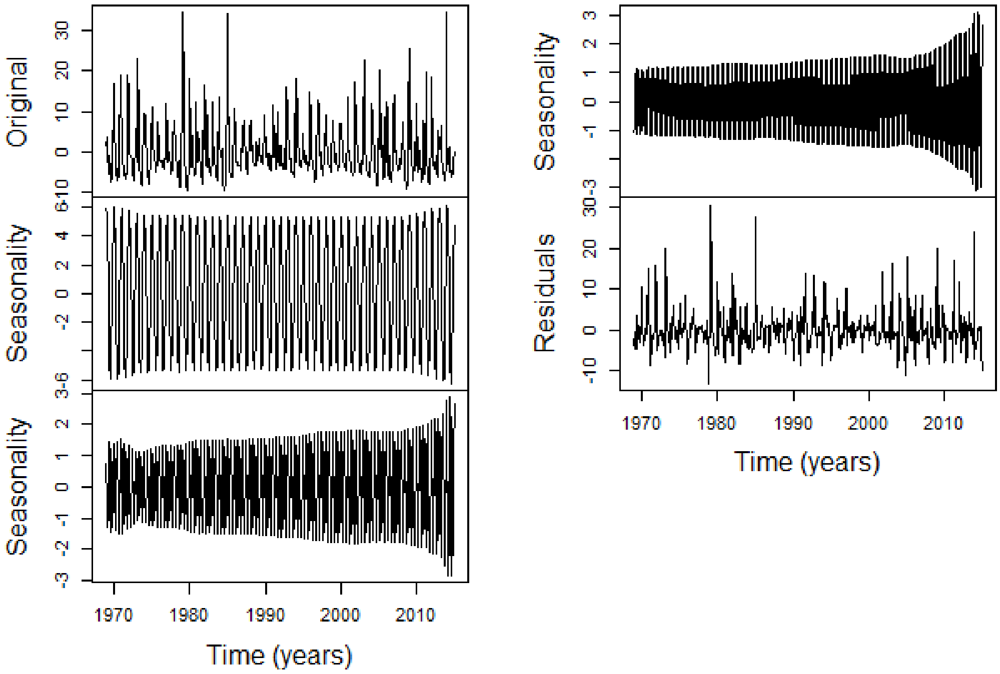

Figure 10. Reconstruction of the seasonality of the Jucu River series.

Figure 11 shows the reconstruction of the original series after grouping the trend component and the seasonal components. The reconstructed flow series variations (in red) follow those of the original flow series (in black), as the maximum/minimum values and the rise/fall rates show. Therefore, the extracted components can better describe the hidden pattern and properties of the original time series (Pham *et al.*, 2021).

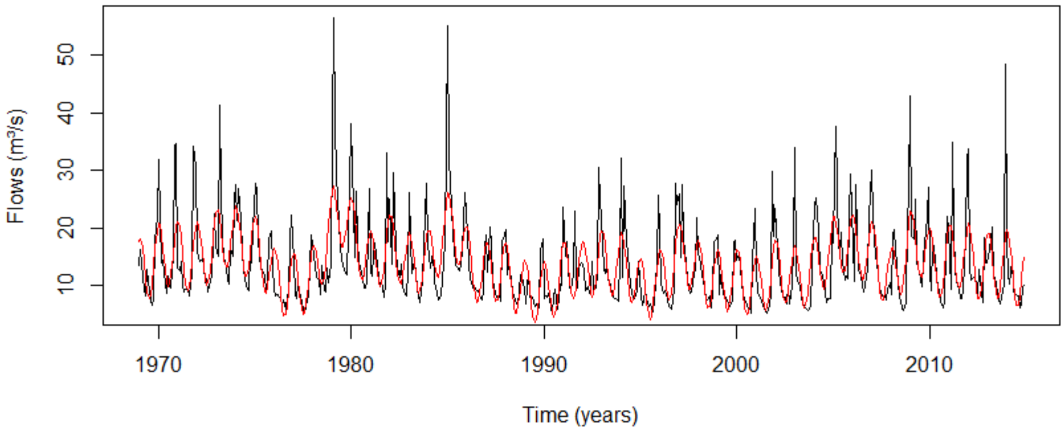

Figure 11. Visual analysis of the original (black) and reconstructed (red) series.

Table 2 presents the performance measures that quantify the quality of the reconstruction. The low values of BIAS, RMSE and MAE and the high values of d_2 and ρ indicate that the reconstruction of the series using the SSA Sequential technique presented good performance and consequently preserved the characteristics of the observed series. Destro *et al.* (2012) and Wang *et al.* (2019b) also verified the good performance of the SSA technique in the reconstruction of a series of the Cuiabá River, Brazil, and of the Gulang River, China, series, respectively. In addition, the high values of BIAS, RMSE and MAE and low values of d_2 and ρ for the noise, indicate the series random characteristic.

Although the use of performance indicators is commonly useful in time series analysis, and can allow understanding the relationship between reconstructed series, noise and observed series, it is important to highlight that limitations can be observed when applying indicators to SSA. As it is a technique that removes noise from the series that invariably presents this component, the possible application of indicators to analyze results coming out from SSA application may not be appropriate. In this context, it is necessary to evaluate how much noise influences the performance of the technique in hybrid or direct modeling applications, comparing SSA with other methodologies, such as SARIMA.

Table 2. Performance measures of series reconstruction.

Subseries	BIAS	RMSE	MAE	d_2	ρ
Reconstruct	-0.0473	0.2206	3.5541	0.8017	0.7855
Noise	13.8940	0.6349	13.894	0.4469	0.3857

Flow is influenced by many factors such as precipitation, temperature, weather conditions and is characterized by non-linear, non-stationary and uncertain conditions (Liu *et al.*, 2019). Due to this, it is common for river flow series to present non-stationary components (Xie *et al.*, 2016), such as seasonality. Analyzing the performance of the series reconstruction, through visual analysis (Figure 11) and quality indicators (Table 2), the reconstruction procedure has an acceptable performance in processing the data homogeneity, allowing to eliminate the inconsistencies of flow series. It is even a viable alternative in the treatment of outliers because, according to Reisen *et al.* (2008), its effects can significantly influence inferences made and, for example, generate errors in forecast analysis and decision-making in production processes, planning and management of water resources.

For example, Pham *et al.* (2021) highlights that the main reason for data pre-processing is to eliminate possible noise in the time series data and, thus, avoid unwanted model training. Chen *et al.* (2021) highlights that useful information hidden in the flow series can be extracted effectively through data pre-processing, thus improving prediction accuracy. Baydaroğlu *et al.* (2018) found, that hybrid modeling of SSA with support vector regression (SVR), on monthly flow data from the Kızılırmak River in Turkey proved successful, yielding values of coefficient of determination (0. 77) and appreciable mean absolute error (0.01). Other authors have identified the good performance of the hybrid SSA to other models in performing flow forecasts, as in Zhang *et al.* (2011), Zubaidi *et al.* (2018) and Apaydin *et al.* (2021).

The identification of periodicities in hydrological data is important for the management of water resources in hydrographic basins and for the establishment of flood forecasting systems. However, the complexity and randomness of hydrological systems make it difficult to detect hidden oscillatory features (Chen *et al.*, 2021; Pham *et al.*, 2021). Therefore, decomposition techniques, such as SSA, are important tools for helping watercourse hydrological behavior understanding. Other studies also verified the good performance of the SSA technique as a time series analysis tool. For example, Marques *et al.* (2006) used SSA to extract components of interest in three variables, among them, flow data, making it possible to verify the ability of SSA to extract a slowly varying component (i.e, the trend) and the oscillatory components of the time series. SSA was also able to accurately predict the components extracted from these time series.

Destro *et al.* (2012) applied the SSA to analyze the behavior of the average monthly flow values recorded at the Acorizal and the Cuiabá fluviometric stations, located on the Cuiabá River, Brazil. The results show that the method is able to extract the trend, harmonics and noise components from the time series. Zubaidi *et al.* (2018) verified the technique's good performance in extracting components of the variable municipal water consumption and six other sets of meteorological data, in a city located in Australia. The results revealed that SSA is a robust technique, capable of decomposing the original time series into many independent components.

4. CONCLUSIONS

This study aimed to apply the SSA technique to average monthly flow data, considering a case study related to the Jucu River Basin, located in Southeastern Brazil. It was possible to observe that the referred series presents a complex behavior and that, as a result, the basic SSA technique would not be adequate to decompose the series. Therefore, the Sequential SSA was applied, which proved to be a successful alternative in terms of series signal and noise components separation quality.

Additionally, it was observed that the technique is a powerful tool for extraction of noise, which is an undesirable component in numerous practical water resources engineering applications, such as forecasting. The tool is powerful for understanding the components of time series and the study addresses important new insights in the time series approach, being a useful tool to help the management and control of water resources.

5. REFERENCES

ABDOLLAHZADE, M.; MIRANIAN, A.; HASSANI, H.; IRANMANESH, H. 2015. A new hybrid enhanced local linear neuro-fuzzy model based on the optimized singular spectrum analysis and its application for nonlinear and chaotic time series forecasting. **Information Sciences**, v. 295, p. 107-125, 2015. https://doi.org/10.1016/j.ins.2014.09.002

ABUDU, S.; CUI, C.; KING, J.P.; ABUDUKADEER, K. Comparison of performance of statistical models in forecasting monthly streamflow of Kizil River, China. **Water Science and Engineering**, v. 3, n. 3, p. 269-281, 2010. https://doi.org/10.3882/j.issn.1674-2370.2010.03.003

AGÊNCIA NACIONAL DE ÁGUAS (Brasil). **HIDROWEB: Sistema de Informações Hidrológicas**. Brasília, 2018. Available at: http://hidroweb.ana.gov.br/default.asp. Access: Apr. 14, 2022.

APAYDIN, H.; TAGHI SATTARI, M.; FALSAFIAN, K.; PRASAD, R. Artificial intelligence modelling integrated with Singular Spectral analysis and Seasonal-Trend decomposition using Loess approaches for streamflow predictions. **Journal of Hydrology**, v. 600, p. 126506, 2021. https://doi.org/10.1016/j.jhydrol.2021.126506

ASWATHAIAH, U.; NANDAGIRI, L. Extraction of nonlinear trends in time series of rainfall using singular spectrum analysis. **Journal of Hydrologic Engineering**, v. 25, n. 12, p. 04020053, 2020. https://doi.org/10.1061/(ASCE)HE.1943-5584.0002017

BALLOT, C.H.D.B. **Les changements périodiques de température, dépendants de la nature du soleil et de la lune, mis en rapport avec le pronostic du temps, déduits d'observations néerlandaises de 1729 à 1846**. Utrecht: Kemink, 1847.

BAYDAROĞLU, Ö.; KOÇAK, K.; DURAN, K. River flow prediction using hybrid models of support vector regression with the wavelet transform, singular spectrum analysis and chaotic approach. **Meteorology and Atmospheric Physics**, v. 130, p. 349-359, 2018. https://doi.org/10.1007/s00703-017-0518-9

BAYER, D.; CASTRO, N.; BAYER, F. Modelagem e previsão de vazões médias mensais do rio Potiribu utilizando modelos de séries temporais. **Revista Brasileira de Recursos Hídricos.** v. 17, n. 2, p. 229-239, 2012. 10.21168/rbrh.v17n2.p229-239

BAYER, F.M.; SOUZA, A.M. Wavelets e modelos tradicionais de previsão: um estudo comparativo. **Revista Brasileira de Biometria**, v. 28, n. 2, p. 40-61, 2010.

BLEIDORN, M.T.; PINTO, W.D.P.; BRAUM, E.S.; LIMA, G.B.; MONTEBELLER, C.A. 2019. Modelagem e previsão de vazões médias mensais do rio Jucu, ES, utilizando o modelo SARIMA. **IRRIGA**, v. 24, n. 2, p. 320-335, 2019. https://doi.org/10.15809/irriga.2019v24n2p320-335

BLEIDORN, M.T.; SCHMIDT, I. M.; KNAAK, J.; PINTO, W. P. Hydrological methodologies analysis to obtaining ecological and grantable flows, considering annual and monthly criteria for Doce River, Brazil. **Ciência e Natura**, v. 45, p. e7-e7, 2023. https://doi.org/10.5902/2179460X68906

BRASIL. Presidência da República. Lei n. 9.433, de 9 de janeiro de 1997. Institui a Política Nacional de Recursos Hídricos, cria o Sistema Nacional de Gerenciamento de Recursos Hídricos, regulamenta o inciso XIX do art. 21 da Constituição Federal, e altera o art. 1º da Lei nº 8.001, de 13 de março de 1990, que modificou a Lei nº 7.990, de 28 de dezembro de 1989. **Diário Oficial [da] União**: seção 1, Brasília, DF, 09 jan. 1997.

CHEN, X.; WANG, X.; LIAN, J. Applicability Study of Hydrological Period Identification Methods: Application to Huayuankou and Lijin in the Yellow River Basin, China. **Water**, v. 13, n. 9, p. 1265, 2021. https://doi.org/10.3390/w13091265

DESTRO, C.A.M.; LIMA, G.A.R.; ZEILHOFER, P. Análise de séries temporais de vazão média mensal do Rio Cuiabá através do método de Análise de Espectro Singular. **Revista Brasileira de Recursos Hídricos**, v. 17, n. 2, p. 111-120, 2012. 10.21168/rbrh.v17n2.p111-120

DU, K.; ZHAO, Y.; LEI, J. The incorrect usage of singular spectral analysis and discrete wavelet transform in hybrid models to predict hydrological time series. **Journal of Hydrology**, v. 552, p. 44-51, 2017. https://doi.org/10.1016/j.jhydrol.2017.06.019

GOLYANDINA, N.; KOROBEYNIKOV, A. Basic singular spectrum analysis and forecasting with R. **Computational Statistics & Data Analysis**, v. 71, p. 934-954, 2014. https://doi.org/10.1016/j.csda.2013.04.009

GOLYANDINA, N.; KOROBEYNIKOV, A.; ZHIGLJAVSKY, A. **Singular spectrum analysis with R**. Berlin: Springer Berlin Heidelberg, 2018.

GOLYANDINA, N.; NEKRUTKIN, V.; ZHIGLJAVSKY, A. A. **Analysis of time series structure: SSA and related techniques**. Boca Raton: CRC press, 2001.

GOLYANDINA, N.; ZHIGLJAVSKY, A. **Singular Spectrum Analysis for Time Series.** Berlin: Springer Berlin Heidelberg, 2013.

HASSANI, H. Singular Spectrum Analysis: Methodology and Comparison. **Journal of Data Science**, v. 5, n. 2, p. 239-257, 2007. 10.6339/JDS.2007.05(2).396

HASSANI, H.; HERAVI, S.; ZHIGLJAVSKY, A. Forecasting UK industrial production with multivariate singular spectrum analysis. **Journal of Forecasting**, v. 32, n. 5, p. 395-408, 2013. https://doi.org/10.1002/for.2244

HASSANI, H.; MAHMOUDVAND, R. **Singular spectrum analysis: Using R**. Springer, 2018.

HASSANI, H.; ZHIGLJAVSKY, A. Singular spectrum analysis: methodology and application to economics data. **Journal of Systems Science and Complexity**, v. 22, p. 372-394, 2009. https://doi.org/10.1007/s11424-009-9171-9

KOMORNÍK, J. *et al*. Comparison of forecasting performance of nonlinear models of hydrological time series. **Physics and Chemistry of the Earth, Parts A/B/C**, v. 31, n. 18, p. 1127-1145, 2006. https://doi.org/10.1016/j.pce.2006.05.006

LIU, S.; HUANG, S.; XIE, Y.; WANG, H.; LENG, G.; HUANG, Q. *et al.* Identification of the non-stationarity of floods: changing patterns, causes, and implications. **Water Resources Management**, v. 33, p. 939-953, 2019. https://doi.org/10.1007/s11269-018-2150-y

MACAULAY, F. R. Appendices to" The Smoothing of Time Series". *In*: MACAULAY, F. R. **The Smoothing of Time Series**. NBER, 1931. p. 118-169.

MALEKIAN, A.; CHOUBIN, B.; LIU, J.; SAJEDI-HOSSEINI, F. Development of a new integrated framework for improved rainfall-runoff modeling under climate variability and human activities. **Water Resources Management**, v. 33, p. 2501-2515, 2019. https://doi.org/10.1007/s11269-019-02281-0

MARQUES, C. A. F.; FERREIRA, J. A.; ROCHA, A.; CASTANHEIRA, J. M.; MELO-GONÇALVES, P.; VAZ, N. *et al.* Singular spectrum analysis and forecasting of hydrological time series. **Physics and Chemistry of the Earth, Parts A/B/C**, v. 31, n. 18, p. 1172-1179, 2006. https://doi.org/10.1016/j.pce.2006.02.061

MENG, E.; HUANG, S.; HUANG, Q.; FANG, W.; WU, L.; WANG, L. A robust method for non-stationary streamflow prediction based on improved EMD-SVM model. **Journal of hydrology**, v. 568, p. 462-478, 2019. https://doi.org/10.1016/j.jhydrol.2018.11.015

MORETTIN, P. A.; TOLOI, C. M. **Análise de Séries Temporais**. 2. ed. Edgar Blücher, 2006.

PETROPOULOS, F.; APILETTI, D.; ASSIMAKOPOULOS, V. *et al.* Forecasting: theory and practice. **International Journal of Forecasting**, v. 38, n. 3, p. 705-871, 2022. https://doi.org/10.1016/j.ijforecast.2021.11.001

PHAM, Q.B.; YANG, T.-C.; KUO, C.-M.; TSENG, H.-W.; YU, P.-S. Coupling singular spectrum analysis with least square support vector machine to improve accuracy of SPI drought forecasting. **Water Resources Management**, v. 35, p. 847-868, 2021. https://doi.org/10.1007/s11269-020-02746-7

PINTO, W. P.; REISEN, V.A.; MONTE, E.Z. Previsão da concentração de material particulado inalável, na Região da Grande Vitória, ES, Brasil, utilizando o modelo SARIMAX. **Engenharia Sanitária e Ambiental**, v. 23, p. 307-318, 2018. https://doi.org/10.1590/S1413-41522018168758

PINTO, W. P.; LIMA, G.B.; ZANETTI, J.B. Previsão de regimes de vazões médias mensais do rio doce, colatina-espírito santo. **Ciência e Natura**, v. 37, n. 3, p. 1-11, 2015.

PRUSKI, F.F.; BOF, L.H.N.; SILVA, L.M.C.; SILVA, J.M. A.; REGO, F.S.; JUSTINO, F.B. Impact of the substitution of reference annual streamflow by monthly streamflow on the potential use of water resources. **Engenharia Agrícola**, v. 34, p. 496-509, 2014. https://doi.org/10.1590/S0100-69162014000300013

R DEVELOPMENT CORE TEAM. **R**: A language and environment for statistical computing. Vienna, 2018.

RAN, Q.; ZONG, X.; YE, S.; GAO, J.; HONG, Y. Dominant mechanism for annual maximum flood and sediment events generation in the Yellow River Basin. **Catena**, v. 187, p. 104376, 2020. https://doi.org/10.1016/j.catena.2019.104376

REISEN, V.A.; MOLINARES, F.A.F.; TEIXEIRA, E.C. Modelagem de séries temporais sazonais na presença de outliers: estudo de caso da vazão máxima mensal do rio Jucu, ES, Brasil. **Revista Brasileira de Recursos Hídricos**, v. 13, n. 2, p. 45-53, 2008. 10.21168/rbrh.v13n2.p45-53

REN, K.; HUANG, S.; HUANG, Q.; WANG, H.; LENG, G. Environmental flow assessment considering inter-and intra-annual streamflow variability under the context of non-stationarity. **Water**, v. 10, n. 12, p. 1737, 2018. https://doi.org/10.3390/w10121737

SALAS, J.D.; DELLEUR, J.W.; YEVJEVICH, V.; LANE, W.L. **Applied Modeling of Hydrologic Time Series.** Littleton: Water Resources Publications, 1980.

SILVA, B.M.B.; SILVA, D.D.; MOREIRA, M.C. Influência da sazonalidade das vazões nos critérios de outorga de uso da água: estudo de caso da bacia do rio Paraopeba. **Revista Ambiente & Água**, v. 10, p. 623-634, 2015. https://doi.org/10.4136/ambi-agua.1587.

UNNIKRISHNAN, P.; JOTHIPRAKASH, V. Data-driven multi-time-step ahead daily rainfall forecasting using singular spectrum analysis-based data pre-processing. **Journal of Hydroinformatics**, v. 20, n. 3, p. 645-667, 2018. https://doi.org/10.2166/hydro.2017.029

UNNIKRISHNAN, P.; JOTHIPRAKASH, V. Hybrid SSA-ARIMA-ANN model for forecasting daily rainfall. **Water Resources Management**, v. 34, p. 3609-3623, 2020. https://doi.org/10.1007/s11269-020-02638-w

WANG, B.; JIANG, Q.; JIANG, P. A combined forecasting structure based on the L1 norm: application to the air quality. **Journal of environmental management**, v. 246, p. 299-313, 2019a. https://doi.org/10.1016/j.jenvman.2019.05.124

WANG, L.; LI, X.; MA, C.; BAI, Y. Improving the prediction accuracy of monthly streamflow using a data-driven model based on a double-processing strategy. **Journal of Hydrology**, v. 573, p. 733-745, 2019b. https://doi.org/10.1016/j.jhydrol.2019.03.101

WU, C. L.; CHAU, K. W.; FAN, C. Prediction of rainfall time series using modular artificial neural networks coupled with data-preprocessing techniques. **Journal of Hydrology**, v. 389, n. 1-2, p. 146-167, 2010. https://doi.org/10.1016/j.jhydrol.2010.05.040

XIE, H.; LI, D.; XIONG, L. Exploring the regional variance using ARMA-GARCH models. **Water Resources Management**, v. 30, p. 3507-3518, 2016. https://doi.org/10.1007/s11269-016-1367-x

ZHANG, Q.; WANG, B.-D.; HE, B.; PENG, Y.; REN, M.-L. Singular spectrum analysis and ARIMA hybrid model for annual runoff forecasting. **Water resources management**, v. 25, p. 2683-2703, 2011. https://doi.org/10.1007/s11269-011-9833-y

ZHANG, X.; WANG, J.; ZHANG, K. Short-term electric load forecasting based on singular spectrum analysis and support vector machine optimized by Cuckoo search algorithm. **Electric Power Systems Research**, v. 146, p. 270-285, 2017. https://doi.org/10.1016/j.epsr.2017.01.035

ZUBAIDI, S.L.; DOOLEY, J.; ALKHADDAR, R.M.; ABDELLATIF, M.; AL-BUGHARBEE, H.; ORTEGA-MARTORELL, S. A Novel approach for predicting monthly water demand by combining singular spectrum analysis with neural networks. **Journal of hydrology**, v. 561, p. 136-145, 2018. https://doi.org/10.1016/j.jhydrol.2018.03.047

APPENDIX A – R CODES USED

The code fragment generally used to extract descriptive statistics is shown below. Initially, the time series is arranged in a spreadsheet in “Oldest to Newest” order. The command data=read.table("clipboard",h=T) is responsible for loading data from the worksheet into the programming environment. The *basicStats* code is responsible for calculating descriptive statistics at a 95% significance level.

```
#-------Code to data input and basic stats calculation-----

data=0

data=read.table("clipboard",h=F)#data input

data

basicStats(data, ci=0.95)#Basic series stats
```

The code fragment generally used to extract the time series graph is shown below. First, the vector (variable that contains the data) is transformed into a time series attributing time characteristics (beginning and end of the series) and the sampling frequency that coincides with the characteristic period of the series (12).

```
#----Code to transform the vector data into a time series--

jucu=ts(data,frequency=12,start=c(1969,1),end=c(2014,12))

plot.ts(jucu,xlab="Time (years)",ylab="Flows (m³/s)")
```

Below is presented a fragment of the programming code for extracting the trend eigenvector, graph of the time series with trend fluctuation, extraction of harmonious eigenvectors, graph of extraction of harmonious components and generation of the graph of weighted correlation.

```
#--------------------SEQUENTIAL SSA CODE-----------------

###-----Code for viewing the trend eigenvector-----

s1<-ssa(jucu, L=12)

summary(s1)# Look inside 's1' object to see, what is available

plot(s1, main=" ",xlab="Eigenvectors",ylab="Singular Values")# Plot of eigenvalues

###-----Code for trend extraction-----

res1<-reconstruct(s1, groups = list(1))

trend<-res1$F1

plot(res1, add.residuals = FALSE, plot.type = "single", main=" ", xlab="Time (years)",ylab="Flows (m³/s)",

col = c("black", "red"), lwd = c(1, 2))

res.trend <- residuals(res1)

###----- Code for viewing harmonious eigenvectors—

s2<-ssa(res.trend,L=276)

plot(s2, main=" ", xlab="Eigenvectors", ylab="Singular Values")

plot(s2, main=" ", type = "paired", idx = 1:12,

plot.contrib = FALSE)

###----- Code to generate the w-correlation graph----

plot(wcor(s2, groups = 1:54), grid = c(7),lwd=4, lty=1,

scales=list(at=c(4,8,12,16,20,24,28,32,36,40,44,48,52)))
```

The code fragment generally used to rebuild the time series without the noise component is shown below. First, the trend component of the time series is reconstructed, and next, the seasonal component. In the last step, the two rebuilt components are joined, resulting in the time series reconstructed without noise.

```
#------Code for rebuilding the series without noise-------

ai<-reconstruct(s1,groups = list(Trend=c(1)))

plot(ai, main=" ",xlab="Time (years)")

aii<-reconstruct(s2,groups = list(Seasonality=c(1,2), Seasonality=c(3,4), Seasonality=c(5,6)))

plot(aii, main=" ", xlab="Time (years)")

aiii=ai$Trend+aii$Seasonality

plot(jucu, ylab= "Flows (m³/s)", xlab="Time (years)")

lines(aiii,col="red")
```