energies

MDPI

Article

Selection of Temporal Lags for Predicting Riverflow Series from Hydroelectric Plants Using Variable Selection Methods

Hugo Siqueira [1], Mariana Macedo [2], Yara de Souza Tadano [3], Thiago Antonini Alves [4], Sergio L. Stevan Jr. [1], Domingos S. Oliveira Jr. [5], Manoel H.N. Marinho [6], Paulo S.G. de Mattos Neto [5], João F. L. de Oliveira [6], Ivette Luna [7], Marcos de Almeida Leone Filho [8], Leonie Asfora Sarubbo [9,10] and Attilio Converti [11,*]

1 Department of Electronics, Federal University of Technology–Parana (UTFPR), Ponta Grossa (PR) 84017-220, Brazil; hugosiqueira@utfpr.edu.br (H.S.); sstevanjr@utfpr.edu.br (S.L.S.J.)
2 BioComplex Lab, Department of Computer Science, University of Exeter, Exeter EX4 4PY, UK; marianagmmacedo@gmail.com
3 Department of Mathematic, Federal University of Technology–Parana (UTFPR), Ponta Grossa (PR) 84017-220, Brazil; yaratadano@utfpr.edu.br
4 Department of Mechanical Engineering, Federal University of Technology–Parana (UTFPR), Ponta Grossa (PR) 84017-220, Brazil; antonini@utfpr.edu.br
5 Departamento de Sistemas de Computação, Centro de Informática, Universidade Federal de Pernambuco (UFPE), Recife (PE) 50670-901, Brazil; dsosj@cin.ufpe.br (D.S.O.J.); psgmn@cin.ufpe.br (P.S.G.d.M.N.)
6 Polytechnic School of Pernambuco, University of Pernambuco (UPE), Recife (PE) 50720-001, Brazil; marinho75@poli.br (M.H.N.M.); fausto.lorenzato@upe.br (J.F.L.d.O.)
7 Department of Economic Theory, Institute of Economics, State University of Campinas (UNICAMP), Campinas (SP) 13083-857, Brazil; iluna@unicamp.br
8 Venidera Pesquisa e Desenvolvimento, Campinas 13070-173, Brazil; marcos@venidera.com
9 Department of Biotechnology, Catholic University of Pernambuco (UNICAP), Recife (PE) 50050-900, Brazil; leonie.sarubbo@unicap.br
10 Advanced Institute of Technology and Innovation (IATI), Recife (PE) 50070-280, Brazil
11 Department of Civil, Chemical and Environmental Engineering, University of Genoa (UNIGE), 16145 Genoa, Italy
* Correspondence: converti@unige.it; Tel.: +39-010-3352593

Received: 14 July 2020; Accepted: 13 August 2020; Published: 16 August 2020

Abstract: The forecasting of monthly seasonal streamflow time series is an important issue for countries where hydroelectric plants contribute significantly to electric power generation. The main step in the planning of the electric sector's operation is to predict such series to anticipate behaviors and issues. In general, several proposals of the literature focus just on the determination of the best forecasting models. However, the correct selection of input variables is an essential step for the forecasting accuracy, which in a univariate model is given by the lags of the time series to forecast. This task can be solved by variable selection methods since the performance of the predictors is directly related to this stage. In the present study, we investigate the performances of linear and non-linear filters, wrappers, and bio-inspired metaheuristics, totaling ten approaches. The addressed predictors are the extreme learning machine neural networks, representing the non-linear approaches, and the autoregressive linear models, from the Box and Jenkins methodology. The computational results regarding five series from hydroelectric plants indicate that the wrapper methodology is adequate for the non-linear method, and the linear approaches are better adjusted using filters.

Keywords: monthly forecasting; autoregressive model; wrapper; bio-inspired metaheuristics extreme learning machines neural networks

1. Introduction

An essential task for countries where power generation is done using hydroelectric plants is the monthly seasonal streamflow series forecasting. This step is directly related to energetic planning, water availability, and pricing strategies. The last Hydropower Status Report from the International Hydropower Association reported that 4306 TWh of electricity was generated in the world employing hydroelectric plants in 2019, corresponding to a 2.5% annual increase [1]. This amount represents the single most significant contribution from a renewable energy source in history. The report addresses data from 13,000 stations in 150 countries. The leaders in hydropower installed capacity are China (356.40 GW), Brazil (109.06 GW), United States (102.75 GW), and Canada (81.39 GW). These figures also provide an idea about the productive chain regarding such power source. It is clear that the correct water management can bring substantial benefits in avoiding water or money waste [2].

Many researchers have addressed streamflow series tasks for different countries such as China, Canada, Ecuador, Iraq, Mozambique, United States, Serbia, Norway, Turkey, Sri-Lanka, and Brazil [2–10]. It highlights the importance of the problem to the global economy. These series present a particular seasonal component due to the periods of rainfall along the year, being non-stationary series [11,12]. However, most of such studies focus on determining just the best predictor, disregarding some other essential steps of the whole process.

Identifying a system is a task influenced by factors, such as prior knowledge of its characteristics, complexity, presence of noise, and performance metrics to be used [13,14]. The inputs must represent the dynamics of the system, which helps in choosing a forecasting model that is appropriate to the problem and is a fundamental step to obtain an efficient model for time series forecasting [15]. The structure of models also defined by the number of entries. For example, in the case of neural networks, the number of inputs impacts the determination of their structure, since the more entries there are, the more complex the neural network will be, as well as the more costly its training, without a guarantee of a performance improvement [16]. Additionally, the number of inputs influences the surface of the cost function, which tends to have more local minima [17,18].

Among potential benefits of inputs selection, we can mention the facilitation of visualization, data understanding, order reduction, memory requirements, reduction in training time, and computational effort [17].

Variable selection (VS) methods attempt to identify a subset of inputs that assist in forecasting, pattern recognition, and data regression, playing a significant role in the accuracy of the forecasting methods. Additionally, such approaches tend to simplify the final model, as well as to improve the stability of responses, and eliminate redundant inputs [19]. Harrell [20] stated that VS could be subjective because they use a conceptual understanding of the dependent variable to select independent variables. Many times, a small number of inputs is recommended for prediction purposes [21]. Guyon and Elisseeff [17] classified the selection methods as filters, wrappers, and embedded. We can also mention a new category, the bio-inspired metaheuristics for optimization.

Many research fields have addressed the importance of variable selection to improve the accuracy of models in different contexts, as presented in the works below:

- Li et al. [22]—to increase the estimation of growing stem volume of pine using optical images;
- Bonah et al. [23]—to quantitative tracking of foodborne pathogens;
- Xiong et al. [24]—to increase near-infrared spectroscopy quality;
- Speiser et al. [25]—an extensive investigation with 311 datasets to compare several random forest VS methods for classification;

- Rendall et al. [26]—extensive comparison of large scale data driven prediction methods based on VS and machine learning;
- Marcjasz et al. [27]—to electricity price forecasting;
- Santi et al. [28]—to predict mathematics scores of students;
- Karim et al. [29]—to predict post-operative outcomes of cardiac surgery patients;
- Kim and Kang [30]—to faulty wafer detection in semiconductor manufacturing;
- Furmańczyk and Rejchel [31]—to high-dimensional binary classification problems;
- Fouad and Loáiciga [5]—to predict percentile flows using inflow duration curve and regression models;
- Ata Tutkun and Kayhan Atilgan [32]—investigated VS models in Cox regression, a multivariate model;
- Mehmood et al. [33]—compared several VS approaches in partial least-squares regression tasks;
- McGee and Yaffee [34]—provided a study on short multivariate time series and many variations of Least Absolute Shrinkage and Selection Operator (LASSO) for VS;
- Seo [35]—discussed the VS problem together with outlier detection, due to each input affecting the regression task;
- Dong et al. [36]—to wind power generation prediction;
- Sigauke et al. [37]—presented a probabilistic hourly load forecasting framework based on additive quantile regression models;
- Wang et al. [38]—to short-term wind speed forecasting;
- Taormina and Chau [39]—to rainfall-runoff modeling;
- Taormina et al. [40]—to river flow forecasting;
- Cui and Jiang [41]—to chaotic time series prediction;
- Silva et al. [42]—to predict the price of sugarcane derivatives;
- Siqueira et al. [12,43–45]—applied the partial autocorrelation function linear filter to streamflow series forecasting;
- Siqueira et al. [2]—use of VS methods, such as wrappers and filters to predict streamflow series; and
- Kachba et al. [46]—application of wrapper and non-linear filters to estimate the impact of air pollution on human health.

Despite these methodologies not being universal or equally useful in various fields, the presentation of the studies above depicts the importance of variable selection for data processing in many different contexts, methods, and tasks. In streamflow series forecasting from hydroelectric plants, this is even more relevant due to the high magnitude of the energy generated. An increase of a single percentage point in the accuracy of such predictions represents an enormous amount of electric power. However, most of the literature focuses on the definition of the best forecasting model, neglecting a further investigation on the impact in adopting distinct VS approaches.

To fill this gap, we analyzed the use of ten VS methods in the autoregressive (AR) model from the Box and Jenkins methodology and the extreme learning machines neural network. The addressed VS approaches are linear filters (two manners of using the partial autocorrelation function, PACF), non-linear filters (three mutual information-based methods), wrappers (considering three evaluating metrics), and bio-inspired metaheuristics (genetic algorithm and particle swarm optimization).

The remainder of this study is organized as follows: Section 2 presents the main content of the variable selection procedure, as well as the main stages of the seasonal streamflow series forecasting; Section 3 discusses the filters; Section 4 the wrapper; Section 5 the bio-inspired metaheuristics—genetic algorithm and particle swarm optimization (PSO); Section 6 the case study, computational results, and the performance analysis; and Section 7 presents the conclusions.

2. Variable Selection

The variable selection methodologies can use information available a priori, through empirical tests of trial and error, or some information criterion. Puma-Villanueva et al. [47] describe a simple example of how the general process works. Consider set **V** that represents the space of input variables, here limited to 3. Thus, we define the vector of inputs $\mathbf{V} = [v_1, v_2, v_3]$, with which it is possible to form $(2^3 - 1 = 7)$ subsets of inputs, as depicted in Table 1.

Table 1. Possible subsets for the input vector **V**.

Subsets	Selected Inputs
1	v_1
2	v_2
3	v_3
4	v_1, v_2
5	v_1, v_3
6	v_2, v_3
7	v_1, v_2, v_3

The selection methods' role is to define which of these subsets is the most appropriate to represent the information in the data, possibly in contrast to the adoption of all inputs. In this case, selecting variables is choosing the subset that allows the best forecast of future values of a time series, that is, selecting the vector $\mathbf{V} \in \Re^k$ among the possible combinations between the variables of $\mathbf{V} \in \Re^l$, such that $k \leq l$. This set represents the dependence structure of a stochastic process over time. In Table 1, the goal of the VS methods is to choose one of seven possibilities.

Yu and Liu [48] present some criteria that relate to this procedure:

- Relevance: the concept associated with the importance of a given variable may have to the problem, since the information it contains will be the basis of the selection process. The relevance is strong or weak depending on how much its removal degrades the performance of the predictor;
- Redundancy: two or more variables are redundant if their observed values are highly correlated or dependent. The level of this correlation reveals the degree of redundancy; and
- Optimality: a so-called optimal subset of input variables is when there is no other subset that produces better results.

However, these characteristics, if combined, have no direct implication. For example, a relevant variable does not mean that the optimal subset contains it. Likewise, the inputs that belong to the optimal subset are not necessarily appropriate [47]. Guyon and Elisseeff [17] classify the VS models into embedded, wrappers, and filters, each of them having its own advantages. The particularities for a given problem indicate which method is most appropriate.

It is essential to point out a difference between the variables and feature selection. The understanding of feature is linked to the idea of a set of inputs that is formed from a combination of the original variables or the extraction of some essential characteristics. An example is principal component analysis (PCA), which linearly combines the inputs [49]. Thus, there is a set of new variables, which are in a new space. It differs from as, in this case, the subset is formed by those variables of the original entries that do not undergo any type of transformation.

Variable Selection in Streamflow Series Forecasting

The predictions regarding the monthly seasonal streamflow series follow the stages shown in Figure 1 (adapted from [2]):

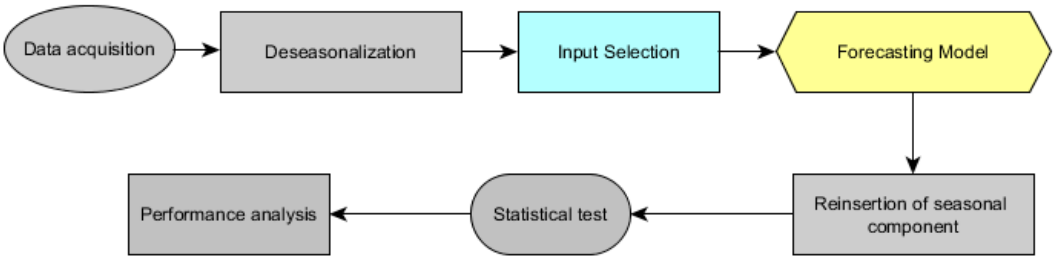


Figure 1. Block diagram of monthly streamflow series prediction.

The first stage is data acquisition. Sensors are spread to measure the volume of water that passes a transversal section of the rivers. Such water is that which moves the turbines of hydroelectric plants. Due to the nature of the rivers, containing many recesses along their way, the measurement's uncertainty incidence is inevitable.

The second stage is the pre-processing, which can be summarized in two steps: (a) the application of the deseasonalization procedure to remove the inherent seasonal component present in this kind of series. The transformed data are stationary series, with zero mean and variance equal to one. For linear models of the Box and Jenkins methodology, this step is mandatory, but some investigations have shown that the performance of non-linear methods also increases with this procedure; (b) the input selection step, to determine the lags that lead to the best performances of the predictors. The input selection step is the focus of this investigation.

The next stage is the definition of the forecasting model. Undoubtedly, this is the most usual theme addressed in streamflow series forecasting. Clearly, the definition of the adequate predictor is crucial regarding the accuracy of the estimation of future samples.

The last stage is post-processing, which involves three procedures. After making predictions, the output responses of the predictors are in the deseasonalized domain. Therefore, we reverse the deseasonalization to allow a performance evaluation in the real domain. Then, a statistical test is applied to verify if the results found by the various predictors are distinct, even if they present different numeric values. Finally, the last step is performance analysis.

Despite the specialized literature on this series is related to the forecasting model, it is necessary to elaborate a specific investigation focused on the input selection process due to the impact it presents in the performance of such models. Determining the most suitable set of lags may lead to distinct conclusions. Thus, this theme must be discussed.

3. Filters

The filter selection method is based only on the available data and does not depend on the predictor model. The variables are chosen through linear or non-linear correlation measures between the observations. The main advantage of this method is its generality, as it is not necessary to synthesize the predictor, which tends to make it computationally efficient [17].

However, as this is a previous step, the optimal set of inputs may not be selected, since there is no interaction with the forecasting model. Metrics based on dependency between samples can be useful, but insufficient to ensure that the chosen set is the best possible. Therefore, we recommend using it for problems with a large amount of available data because if the criterion of optimality is not met, the computational cost should be worthwhile.

Figure 2 shows the scheme of the filter type method. Some of the inputs contained in the vector $\mathbf{u}_t$ will belong to vector $\mathbf{u}'_t$ of smaller or equal dimension. The predictions are performed with $\mathbf{u}'_t$.

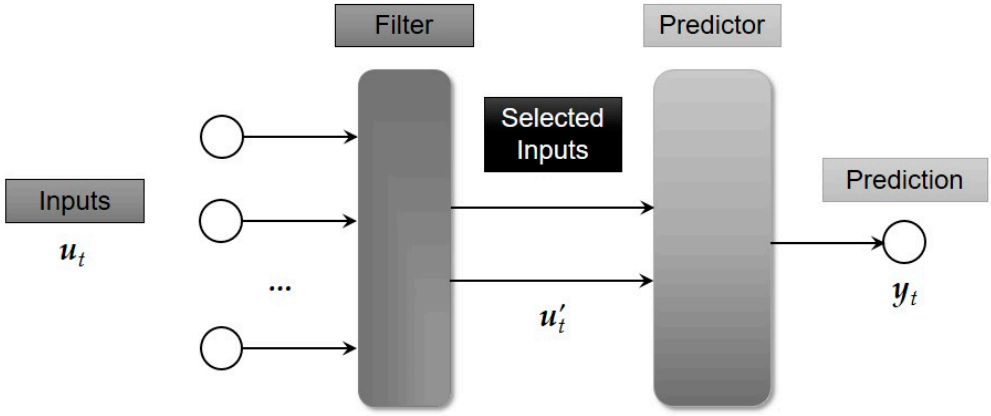


Figure 2. Schematic of the filter model for variable selection.

3.1. Partial Autocorrelation Function

The partial autocorrelation function (PACF) is a widely used filter to identify the order of linear models [11]. The definition of the partial correlation coefficient is directly related to autoregressive models (AR) and the Yule–Walker equations.

The partial autocorrelation coefficient of order k is the last coefficient of an AR(k) model, adjusted for a time series $\mathbf{x}_t$ and denoted by φ_{kk}. This means that an AR process of order p is different from zero to k less than or equal to p, and zero for $k > p$. Based on this assumption and using the Yule–Walker equations, the relationship between the autocorrelation estimates of a time series in these terms obeys the set of equations described in (1):

$$\rho_j = \varphi_{k1}\rho_{j-1} + \varphi_{k2}\rho_{j-2} + \ldots + \varphi_{k(k-1)}\rho_{j-(k+1)} + \varphi_{kk}\rho_{j-k}, j = 1, 2, \ldots, k \quad (1)$$

or, in a matrix form, we have (2) and (3):

$$\mathbf{P}_{kp} = \begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{p-1} \\ \rho_1 & 1 & \cdots & \rho_{p-2} \\ \vdots & \vdots & & \vdots \\ \rho_{p-1} & \rho_{p-2} & \cdots & 1 \end{bmatrix}, \boldsymbol{\rho}_{kp} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{bmatrix}, \boldsymbol{\Phi}_{kp} = \begin{bmatrix} \varphi_{k1} \\ \varphi_{k2} \\ \vdots \\ \varphi_{kp} \end{bmatrix}, \quad (2)$$

$$\boldsymbol{\Phi}_{kp} = \mathbf{P}_{kp}^{-1}\rho_{kp}. \quad (3)$$

in which ρ are the coefficients of the autocorrelation [50].

Thus, the AR(p) model of order p = 1, 2, ..., k must be adjusted to find φ_{kk}. Expression (4) shows the coefficients of the first two AR models:

AR(1): $\rho_1 = \varphi_{11}\rho_0$

$$\mathbf{P}_{1p} = [\rho_0],\ \boldsymbol{\rho}_{1p} = [\rho_1],\ \boldsymbol{\Phi}_{1p} = [\varphi_{11}], \quad (4)$$

being $\varphi_{11} = \rho_1$ the first partial autocorrelation coefficient. Similarly, we have (5):

AR(2): $\begin{vmatrix} \rho_1 = \varphi_{11}\rho_0 + \varphi_{22}\rho_1 \\ \rho_2 = \varphi_{21}\rho_1 + \varphi_{22}\rho_0 \end{vmatrix}$

$$\mathbf{P}_{1p} = \begin{bmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{bmatrix}, \boldsymbol{\rho}_{1p} = \begin{bmatrix} \rho_1 \\ \rho_2 \end{bmatrix}, \boldsymbol{\Phi}_{1p} = \begin{bmatrix} \varphi_{21} \\ \varphi_{22} \end{bmatrix}, \quad (5)$$

Isolating φ_{21} and equating the equations, we have $\varphi_{22} = \left(\rho_2 - \rho_1^2\right)/\left(1 - \rho_1^2\right)$, the value admitted as the second partial autocorrelation coefficient.

It is noteworthy that the autocorrelation coefficients ρ_p are problem-dependent. Then, the PACF of a series can be estimated through successive adjustments of the autoregressive models, determining

the most appropriate orders of an AR model. In practice, the AR(1) is adjusted, from where we estimate the coefficient φ_{11}. Following, we adjust the AR(2), and we have φ_{21} and φ_{22}, the latter being of interest. We continue in these systematic steps until the required k order is adjusted, from where the coefficients come out the desired φ_{kk}.

For a time series, the highest order is sought such that all estimates φ_{kk} for $k > p$ are not significant. The order of the model is the value corresponding to the selected entry, that is, if the coefficients are selected φ_{11} and φ_{55}, lags 1 and 5 are part of the subset of inputs.

Quenouille [51] showed that for a AR(p) process, the coefficients φ_{kk} estimated for orders greater than p have a Gaussian distribution with a mean equal to zero, variance equal to $VAR[\varphi_{kk}] \cong 1/N$, being N the number of samples. Thus, the confidence threshold for the coefficients based on the standard deviation is $|\varphi_{kk}| \cong 2/\sqrt{N}$, considering that the estimate is different from zero in this interval.

However, the method can select non-consecutive delays as model inputs. For example, if $\mathbf{V} = [v_1, v_2, v_3, v_4]$, it can select $\mathbf{V} = [v_1, v_4]$, which means that φ_{11} and φ_{44} were significant, while φ_{22} and φ_{33} are not. For hydrologic time series, Stedinger [52] states that it makes no sense that a given sample is related to non-consecutive delays and that delays in a non-consecutive hydrological system selected by the PACF have no physical meaning, proposing the suppression of these entries. According to this work, some historical series have an autocorrelation structure relative to both the time between observations and the observed period.

Taking as an example an AR(6) model, if PACF defines that only the inputs weighted by the coefficients φ_{11} and φ_{44} are significant, the latter is considered as spurious data and must be discarded. This means that intermediate values should not be considered. Siqueira et al. [2] used bootstrapping techniques to evaluate the best order of periodic autoregressive models, and reached the same conclusion as Stedinger, with similar orders for streamflow series.

Figure 3 is related to the calculation of the PACF of the monthly streamflow series from Furnas hydroelectric plant, located in Brazil. The data used are from January. The horizontal line is the confidence threshold calculated as a function of the standard deviation. Note that, with 12 delays, the method selected lags 1, 5, and 7. If Stedinger's proposal is taken into account, only delay 1 is chosen.

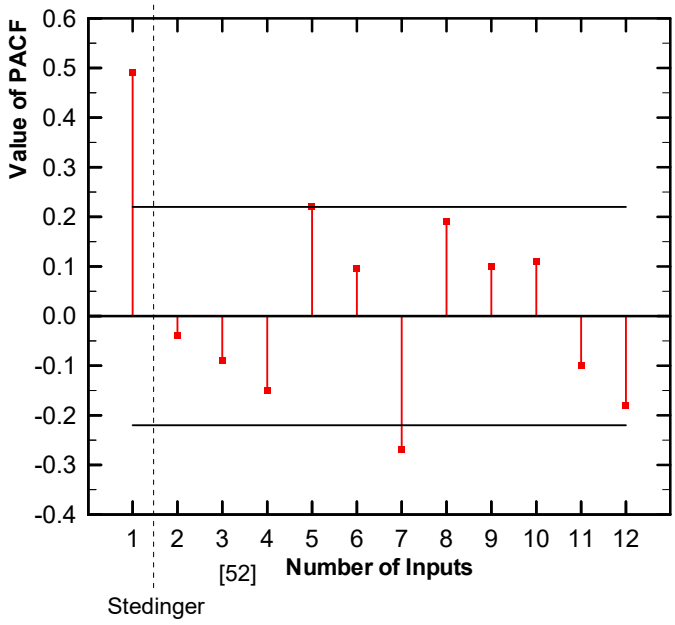


Figure 3. Example of partial autocorrelation function.

The technique is implemented for the selection of inputs in linear simulation models by the Brazilian National Electric System Operator [12].

3.2. Mutual Information

The dependency between two variables is an essential step for selecting model inputs. In this context, a reliable criterion belonging to the scope of information theory can be used, i.e. The mutual information (MI) [53]. The MI is a metric that provides a measure of the degree of dependence between variables; it reflects the amount of information that links them. This filter can be applied as a criterion to select the inputs of forecasting models.

The definition of MI between two random variables can be interpreted as a measure of proximity between the joint probability distribution of the variables x and y, and the product of their marginal distributions. Mathematically, we have (6):

$$MI = \int f_{xy}(x,y)\ \log\left(\frac{f_{xy}(x,y)}{f_x(x)\,f_y(y)}\right) dx\,dy, \tag{6}$$

in which $f_{xy}(x,y)$ is the joint probability density function (PDF), and $f_x(x)$ and $f_y(y)$ are the respective marginal density functions. The MI criterion presents zero as a result for independent variables, and greater than zero otherwise. If a representative sample of the data is available, one can estimate (6) using (7) [54]:

$$MI = \frac{1}{N}\sum_{i=1}^{N} \log\left[\frac{f_{xy}(x_i,y_i)}{f_x(x_i)\,f_y(y_i)}\right], \tag{7}$$

where (x_i, y_i) is the i-th pair of data from the sample with size N, being i = 1, 2,..., N.

The difficulty in this case is to estimate the probabilities since the distributions are often unknown in practice. Additionally, these estimates may require a large amount of data, which is not always available.

There are several ways to estimate PDFs in the literature. In this work, we use a non-parametric approach based on kernel functions, the absolute distance, or city-block type, which have already been applied in streamflow series forecasting [54]. The choice for this proposal is justified in terms of computational simplicity and absence of data distribution type assumption.

Consider the input and output dataset $[\mathbf{X}_k, y_k]$, being k = 1, 2,..., N. The approximation of the probability densities of a one-dimensional x variable via non-parametric kernel estimators is given by (8):

$$\hat{f_x} = \frac{1}{N\lambda}\sum_{i=1}^{N} K\left[\frac{x-x_i}{\lambda}\right] = \frac{1}{N}\sum_{i=1}^{N} K_\lambda(x-x_i), \tag{8}$$

in which $K_\lambda(t)$ is the kernel function, and λ the bandwidth or dispersion parameter.

Therefore, the marginal approximate probability density function of $\mathbf{x}$ is given by (9):

$$\hat{f_x}(\mathbf{x}) = \frac{1}{N(2\lambda)^p}\sum_{i=1}^{N} \exp\left[-\frac{1}{\lambda}\sum_{i=1}^{p}\left|\mathbf{x}_j - \mathbf{x}_{ij}\right|\right], \tag{9}$$

with p being the dimension of $\mathbf{x}$.

Equation (9) arises from (8) as a case adapted to multidimensional $\mathbf{x}$, and using the city-block function. The parameter λ is calculated by (10) [55]:

$$\lambda = \left(\frac{4}{p+2}\right)^{\frac{1}{(p+4)}} N^{\frac{-1}{(p+4)}}. \tag{10}$$

Finally, the joint probability of $(\mathbf{x} - y)$, the latter being a one-dimensional output, is as in (11) [56]:

$$\hat{f_{xy}}(\mathbf{x},y) = \frac{1}{N(\lambda)^{p+1}}\sum_{i=1}^{N} K\left(\frac{\mathbf{x}-\mathbf{x}_j}{\lambda}\right) K\left(\frac{y-y_j}{\lambda}\right), \tag{11}$$

or as in (12)

$$\hat{f}_{xy}(\mathbf{x}, y) = \frac{1}{N(\lambda)^{p+1}} \sum_{i=1}^{N} \exp\left(-\frac{1}{\lambda} s_i\right). \tag{12}$$

Therefore, s_i is calculated by (13):

$$s_i = \sum_{j=1}^{p} \left|\mathbf{x}_j - \mathbf{x}_{ij}\right| + \left|y - y_j\right|. \tag{13}$$

An example that shows the approximation capability of this proposal is its use to build a bi-variable Gaussian distribution function. This function is defined by (14):

$$f_{xy} = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp(-\Gamma) \tag{14}$$

where:

$$\Gamma = \frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_x)}{\sigma_x^2} + \frac{\left(y-\mu_y\right)}{\sigma_y^2} + 2\rho\,\frac{(x-\mu_x)\left(y-\mu_y\right)}{\sigma_x^2\,\sigma_y^2}\right]. \tag{15}$$

with x and y being the variables used, μ_x and μ_y their respective averages, σ_x and σ_y the standard deviations, and ρ the correlation coefficient between them.

To exemplify the approximation capability using the city-block function, we generate 2000 samples of x and y with normal distribution and zero mean. With Equation (14), it is possible to plot f_{xy}, represented graphically in Figure 4a together with its diagram in contour lines (Figure 4b). In parallel, we present the approximations using the city-block kernel function of (13) in Figure 4c,d. The proximity of the curves is clear, which illustrates how this function approximates the distributions. The correlation coefficient between them is 0.9932, although the circles are not perfectly concentric.

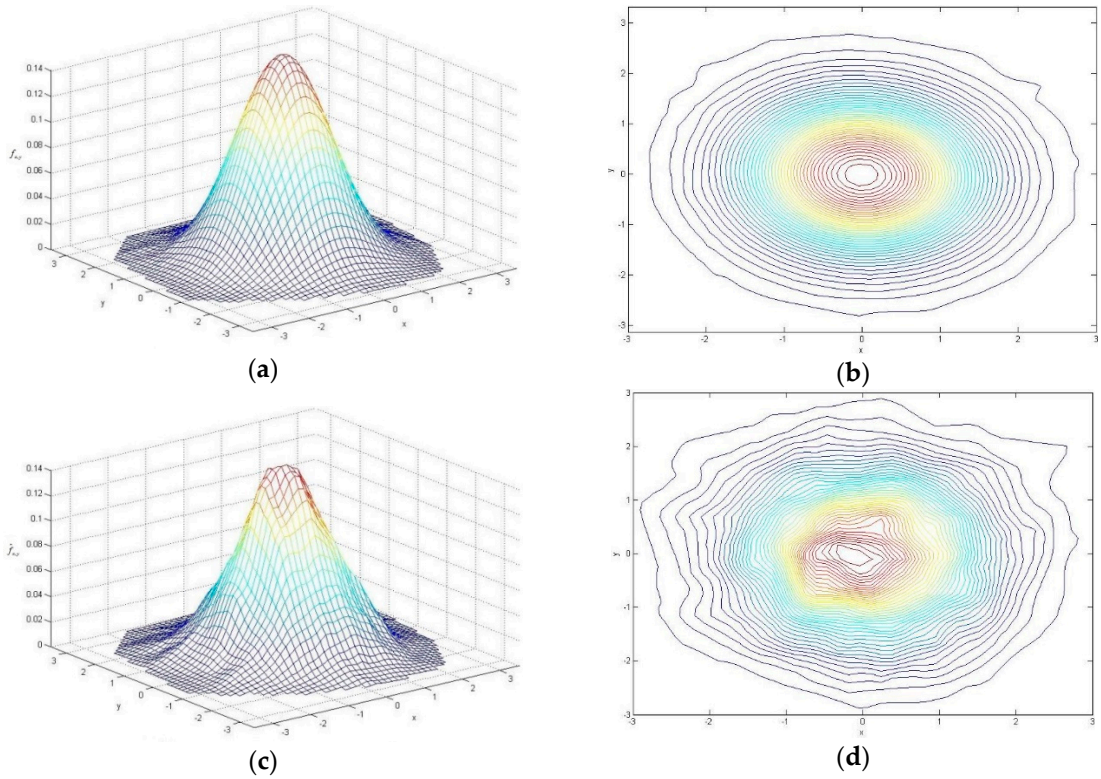


Figure 4. Original bi-variable Gaussian function (**a**) and (**b**), and the approximation (**c**) and (**d**).

The final step in applying the MI filter is to define a confidence threshold that determines if an input belongs to the selected subset of explanatory variables. One possibility is to establish a minimum value for the MI and reject entries with a lower MI score. Another option is to use a bootstrapping or resampling technique to test the hypothesis of independence between x and y. For this, several sequences other than x in relation to y are built, in which the independent variable is reordered, and a vector of MIs is obtained. If this value surpasses the threshold at a given level of significance α, x and y are considered dependent. Thus, an input variable x is identified [57]. This work adopts the latter approach.

The example in Figure 5 refers to the calculation of the MI coefficients related to Furnas hydroelectric plant streamflow series. As in Figure 3, the samples are from January. In this case, we adopted $p = 100$ sequences and $\alpha = 5\%$, and the respective MI values calculated. For 12 lags initially considered as possible explanatory variables, the selected ones are lags 1, 8, and 9.

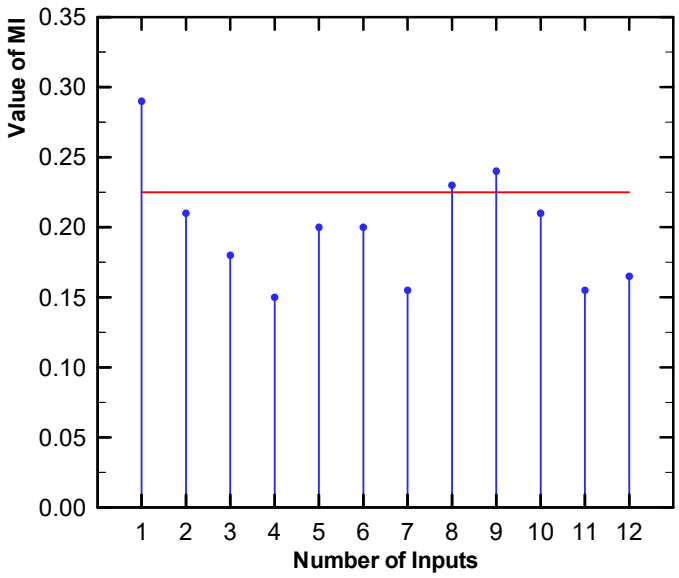


Figure 5. Example of mutual information values.

3.3. Partial Mutual Information

The MI criterion appears advantageous due to the non-parametric nature on identifying explanatory variables, regardless of the model's nature to adjust afterward. Notwithstanding, two or more explanatory variables may be highly correlated; thus the choice of those variables would insert redundancy and an unnecessary increase in the complexity of the model. It may occur because the criterion does not perform a joint evaluation of the whole set of potential input variables. One way to deal with this is the proposal of [58], who reformulates the MI into what is known as the partial mutual information criterion (PMI). The PMI criterion measures the mutual information between the independent variable x and the dependent variable y, conditioned to a set of inputs $\mathbf{z}$ previously selected.

Consider that this set $\mathbf{z}$ exists. Next, it is necessary to extract the influence of this set concerning the other potential inputs evaluated yet, to calculate their real contribution, a different from the one already given by $\mathbf{z}$. Thus, following Luna et al. [54] and Sharma [58], Equation (7) can be reformulated as (16):

$$PMI = \frac{1}{N}\sum_{i=1}^{N} \log_e\left[\frac{f_{x'y'}\left(x'_i, y'_i\right)}{f_{x'}\left(x'_i\right) f_{y'}\left(y'_i\right)}\right] \tag{16}$$

where $x' = x - E(x|\mathbf{z})$ and $y' = y - E(y|\mathbf{z})$.

Here, x' and y' denote the residuals of x and y, respectively, after removing the conditional expectations given **z,** $E(x|\mathbf{z})$ and $E(y|\mathbf{z})$. With this transformation, x' and y' can be interpreted as remaining information in both variables: what is yet different from the information in **z**; and what has not been explained yet from y.

Several approaches have been proposed for estimating expected means [58–60]. We opted for simplicity by following the non-parametric approach based on kernel regressions by using the Nadaraya–Watson estimator [61]. According to this, given two variables a and b, a general expected value $E(a|b)$ is defined by (17):

$$\hat{r}(a) = \sum_{i=1}^{N} w_{\lambda a}(a, a_i)\, b_i \tag{17}$$

where:

$$w_{\lambda a}(a, a_i) = \frac{K_{\lambda a}(a - a_i)}{\sum_{i=1}^{N} K_{\lambda a}(a - a_i)}. \tag{18}$$

with $K_{\lambda a}(a - a_i)$ denoting the kernel function for variable a. As before, we will use the city-block function for this purpose.

Therefore, input selection, in this case, follows an iterative process. At the first iteration, MI scores are calculated for every potential input variable previously defined. The first input selected is the one with the higher MI score as long as its significance is validated, initiating **z**. In the following steps, PMI scores are calculated for all the potential input variables, updating **z** at each iteration, until the higher PMI is not statistically significant at all. The bootstrapping technique is once more used to verify the PMI scores significance at a 5% level.

3.4. Normalization of Maximum Relevance and Minimum Common Redundancy Mutual Information

Some studies use the principles of mutual information, extending it in different directions to increase the filers' selection capability. The work from Che et al. [62] proposed to expand the MI using the maximum relevance and minimum common redundancy (MRMCR) between the inputs of a model (lags). This framework intends to determine the best set of inputs, controlling the redundancy between them.

The first step of this method is to calculate the common redundancy to evaluate the inputs' common information. Following, one must apply the normalization of maximum relevance and minimum common redundancy (N-MRMCR-MI). The result presents values in the interval [0,1].

Let S be the subset of chosen inputs, and T the complementary non-selected group. Then, calculate the common mutual information (CI), using (19):

$$CI(x_i, S, y) = \max_{x_j \in S} \left\{ \frac{MI(x_i, x_j)}{max\left[MI(x_i, x_j), MI(x_i, y), MI(x_j, y)\right]}\, min\left[MI(x_i, y), MI(x_j, y)\right] \right\}, \tag{19}$$

where i is the index of the variables in T, j the index of the inputs in S, and MI is the mutual information (see (7)):

The complete application of N-MRMCR-MI procedure is according to the following stages [62]:

(1) Initialization: be $T = \left(x_1, x_2, \ldots, x_p\right)$ the full set of inputs, and $S = \varnothing$ (empty);

(2) First input selection: calculate $F(x_i)$ using (20) for all i = 1, 2, … , p, and set the best on x_i^* applying (21):

$$F(x_i) = \frac{MI(x_i, y)}{MI(y, y)}, \tag{20}$$

$$x_i^* = \underset{i=1,2,\ldots p}{\operatorname{argmax}}\{F(x_i)\} \tag{21}$$

(3) Update the groups: $T = T - \{x_i^*\}$, and $S = \{x_i^*\}$;
(4) Greedy selection: repeat steps 1 and 2 until the desired number of features is determined;
(5) Determine the N-MRMCR-MI considering the output variable using (22):

$$F(x_i) = \frac{MI(x_i, y)}{MI(y, y)} - \frac{CI(x_i, S, y)}{MI(y, y)}, \tag{22}$$

(6) Update $T = T - \{x_i^*\}$, and $S = \{x_i^*\}$;
(7) Output the subset S.

In this work, we again address the bootstrapping to calculate the confidence level [57], and the city-block functions.

4. Wrappers

In the wrapper approach, the central aspect is the interaction between the variable selection mechanism and the forecasting model [63]. Once the model has already been adjusted, the wrapper will evaluate, through some performance criteria, each of the subsets to solve it [47]. However, the computational cost involved is high, as the model needs to be adjusted for each candidate subset [64]. The literature recommends using this method for cases where the number of samples is reduced [17]. Figure 6 shows the scheme of the method.

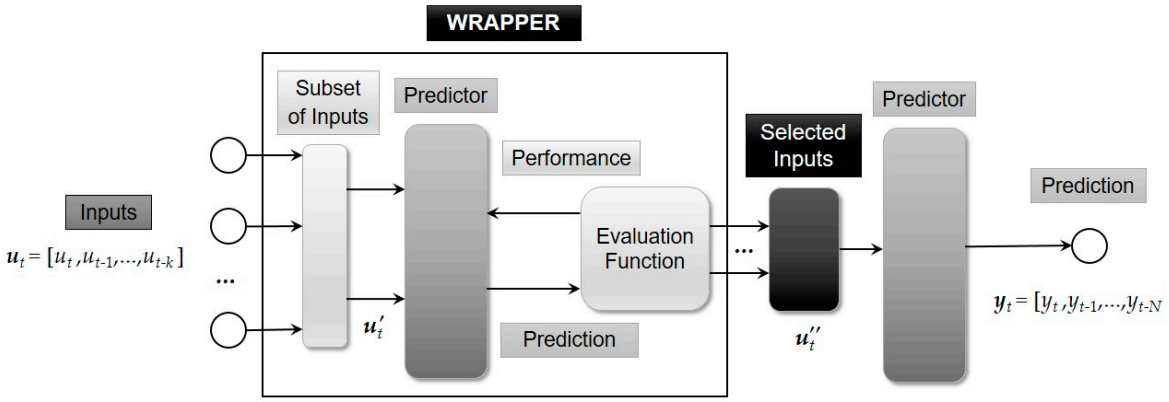


Figure 6. Schematic of the wrapper method.

As shown in Figure 6, the operation of the wrapper is as follows: firstly, the set of entries $\mathbf{u}_t$ is divided into smaller subsets $\mathbf{u}'_t$; next, the predictor is trained and executed for each of subset; after the forecasting stage, we calculate a performance score from the evaluator block for each subset. The one with the best value is used.

It is possible not to use the last predictor shown in Figure 6, if the result of each assessment is stored. However, we presented the selection as a separate task of the forecasting process for the sake of simplicity of understanding.

4.1. Progressive Selection

The computational cost of performing an exhaustive search of all possible subsets can be impractical even a relatively small problem since the computational cost is factorial. A proposal to overcome this problem is the wrapper using the progressive selection method. This methodology establishes a manner to build subsets of entries considering each one individually.

The procedure initiates with an empty subset, and we compare each variable with all others. The one presenting best performance measured by the evaluation function is selected, either to improve the result or to least deterioration of this value. After choosing the first entry, fixed in the subset,

the others are evaluated to ascend as the second entry. We repeat this procedure until the evaluation of all V variables. The final subset is the one with the best overall result.

Figure 7 presents this idea, considering the Emborcação series, with the adjustment of an Extreme learning machines neural network (ELM) with a fixed number of 20 neurons in the hidden layer, and a maximum of 10 delays as input. As one can note, we selected three entries in this order: 4, 10, and 6, since this was the combination that had the lowest mean square error. In this case, it is noticeable that the selected entries are not consecutive and that the increase in the number of inputs does not necessarily improve performance.

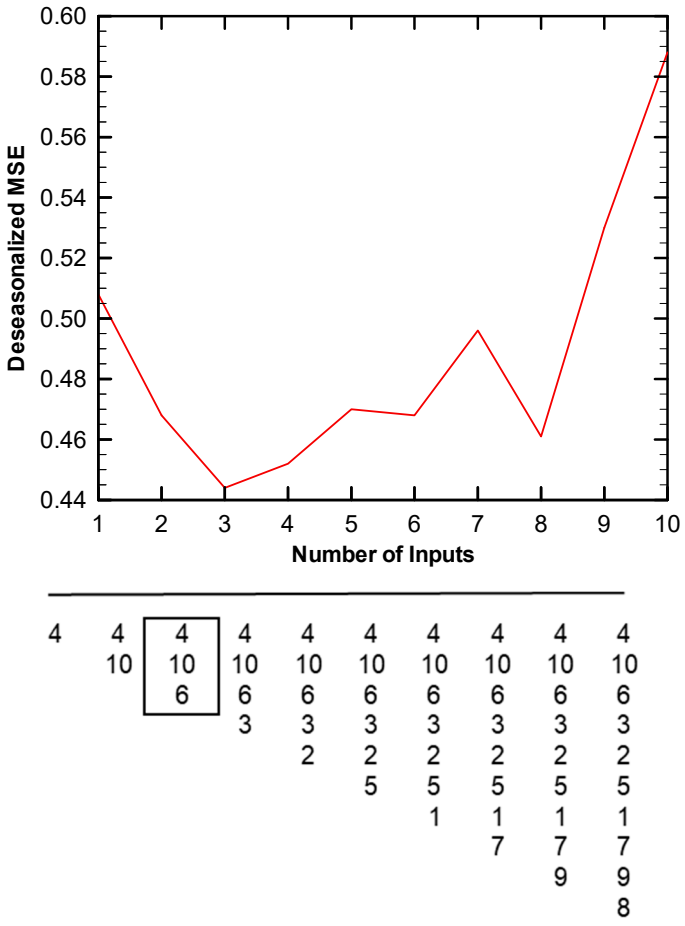


Figure 7. Error behavior with the progressive selection method.

The number of subsets formed in this case is equal to the number V of inputs, and the number of times the predictor needs training obeys $V(V+1)/2$. In the example, above 55 ELMs were adjusted, since $V = 10$.

It is also interesting to observe the behavior of the mean squared error (see Section 4.2) between iterations 2 and 4 in Figure 7. When adding the lag v_6, the error decreased, improving the value of the objective function to be minimized. However, with the insertion of the variable v_3, the error increased. This behavior occurs because the search may fall in local minima, which can circumvent at later iterations [47].

4.2. Evaluation Functions

After discussing how the wrapper method works, it is necessary to define a criterion for assessing the quality of forecasting using the previously defined subsets. This step corresponds to the Evaluator block in Figure 7.

The most straightforward criterion is to use some error metric that, for each adjusted set, shows the average of the differences between the desired and the predicted data. The proposal used by Puma-Villanueva et al. [47] is the mean absolute error (MAE), given by (23):

$$MAE = \frac{1}{N_s}\sum_{t=1}^{N_s}|x_t - \hat{x}_t|, \tag{23}$$

where x_t is the desired sample in time t, $\hat{x}_t$ is the prediction, and N_s the number of predicted samples.

Another possibility is to address the mean squared error (MSE), the most common metric used as a cost function in the training of neural networks, and in the estimation of AR model parameters. This metric is defined by (24):

$$MSE = \frac{1}{N_s}\sum_{t=1}^{N_s}(x_t - \hat{x}_t)^2. \tag{24}$$

Note that these criteria only consider the final result of the adjustment, regardless of the number of inputs. However, there are other types of evaluation functions seek to penalize the number of entries in order to select parsimonious subsets regarding the number of inputs. Criteria widely used are based on information measures [17].

Schwarz [65] proposed the Bayesian Information Criterion (BIC). It is based on linear correlation metrics, and is linked to the optimal orders of the forecasting model, as defined in (25):

$$BIC = N\log_e\left(\hat{\sigma}_a^2\right) + p\log_e(N). \tag{25}$$

where N is the number of observations, p is the order or number of model entries and $\hat{\sigma}_a^2$ is estimated variance of white noise (or residue).

Thus, the wrapper chooses the set of inputs with the lowest BIC value. Similarly, the Akaike Information Criterion (AIC) [66] is defined by (26):

$$AIC = N\ln\left(\hat{\sigma}_a^2\right) + 2p. \tag{26}$$

In both cases, it is clear that there is a penalty concerning the number of entries so that the inclusion of more inputs depends not only on the performance improvement but also on how much it is increased. Thus, the selected subset must be efficient and parsimonious. The difference between the criteria lies in the fact that the BIC penalizes the inclusion more strongly than the AIC. Observe the BIC last term (25) is the natural logarithm of the number of observations, while the AIC (26) has a multiplication of the order or number of model entries by 2.

5. Bio-Inspired Metaheuristics

Bio-inspired metaheuristics for optimization have been widely applied in VS tasks, especially their binary versions. The genetic algorithm (GA) belongs to the field of evolutionary computation because it is inspired by Darwinian natural selection. Another class of bio-inspired methods is the swarm-based approaches, from which many algorithms were created. The main characteristic of this class is the inspiration based on the collective behavior of groups of animals. Its primary representative is the particle swarm optimization (PSO) [67]. Still, many other algorithms can be cited as artificial bee colony, cat swarm optimization, fish school search, ant colony optimization, among others [68,69].

In this section, we briefly explain the two metaheuristics widely used in the literature for optimization: genetic algorithm and particle swarm optimization. Both techniques simulate multiple agents that evolve/adapt depending on the environment to find better solutions to one fitness function. The flexibility, robustness, and scalability are key advantages of applying metaheuristics to real problems.

5.1. Genetic Algorithm

The genetic algorithm (GA) was introduced by John Holland [70], and the theory of natural evolution inspires it. The population which adapts better to the environment circumstances perpetuates the genes, reproducing other individuals with more likely useful genes. Therefore, GA has a population of individuals that will pass into three processes: selection, crossover, and mutation to reproduce a better population until an a priori constraint is reached. For the binary version, a vector of binary values represents each individual of the population.

In summary, the GA starts generating an initial (generally random) population of individuals with the respective fitness function of their genes. Until the population has not converged, the three processes will be executed, and the fitness function will be updated when the genes are updated. The first process, selection, chooses the individuals that will reproduce new individuals. Then, in the following procedure, crossover, the adopted parents (two) will mix their genes to create a new offspring. Thirdly, the mutation process randomly selects genes to be changed. In general, each process establishes the number of individuals that will be updated. The algorithm is usually greedy (only allowing the update of the new individuals that are better than the current ones).

5.2. Particle Swarm Optimization

Kennedy and Eberhart [67] developed particle swarm optimization (PSO). After two years, the discrete or binary version of PSO was also published by them [71]. The PSO algorithm mimics the behavior of a flock of birds where each bird is a candidate solution. Each candidate solution i is represented by a position x_i and a velocity v_i. For the binary version, binary values represent the position and the velocity is continuous values between 0 and 1.

In summary, PSO starts generating an initial (generally random) population of birds with their respective fitness function. Then, until reaching a priori condition, all the birds update the velocity and position. The update of the post (flip the position for binary optimization) is performed each time that random value is higher (or smaller) than a transformation function of the velocity $F(v)$ (such as sigmoid or tangent sigmoid). Using the sigmoid function, we can express that the new position is updated by:

$$x_i^{t+1} = \begin{cases} 0 \, if \; rand() \geq F\left(v_i^{t+1}\right) \\ 1 \, if \; rand() < F\left(v_i^{t+1}\right) \end{cases} \tag{27}$$

Moreover, the new velocity is calculated based on the current velocity, and the delta displacement between the personal ($p^t_{best_i}$) and global ($g^t_{best_i}$) best positions. The $p^t_{best_i}$ is updated every time that a particle finds a better position, and the $g^t_{best_i}$ is the best position between the neighbors' particle (defined a priori by the topology). The parameters c_1 and c_2 are a priori constants that define how altruistic or selfish each particle is, and the parameters r_1 and r_2 are random values between 0 and 1. Equation (28) shows the update process in the velocity:

$$v_i^{t+1} = wx_i^t + c_1 r_1\left(p^t_{best_i} - x_i^t\right) + c_2 r_2\left(g^t_{best_i} - x_i^t\right). \tag{28}$$

Until the algorithm is converged, each particle updates the velocity and position. When the position is updated, the fitness function is also updated.

6. Case Study

In this section, we summarize the computational results of the linear model and the Extreme Learning Machines neural networks using the variable selection techniques discussed: filters, wrappers, and bio-inspired metaheuristics. As discussed in Section 1, seasonal streamflow series forecasting is essential for countries presenting hydroelectric plants to power generation. In the Brazilian case, 70% of the electric energy is hydroelectric generated [72]. Additionally, this task is vital to optimize the energetic planning [43–45,73].

First, it is paramount to mention the adopted assumptions. We performed the simulations with a maximum of 6 delays, following the literature [2,11,54]. We addressed two forecasting approaches: the use of one predictor to the whole series, and the monthly approach, in which we adjusted 12 different models, one for each month of the year.

Several investigations have shown that monthly streamflows present a seasonal behavior throughout the year, being non-stationary series. Linear models cannot be directly applied, being necessary to remove the seasonal component. In this work, we adopted the deseasonalization procedure to transform the series into stationary, with zero mean and variance equals one [2]. The process is reversed before the performance analysis. Equation (29) describes the deseasonalization procedure:

$$z_{i,m} = \frac{s_{i,m} - \hat{\mu}_m}{\hat{\sigma}_m}. \tag{29}$$

where, $\mathbf{s}_n$ is the original series formed by the samples $s_{i,m}$, which is transformed into a new series $\mathbf{z}_n$; $\hat{\mu}_m$ is the monthly mean; $\hat{\sigma}_m$ the monthly standard deviation; and the month $m = 1, 2 \ldots, 12$.

The series addressed are related to five important Brazilian hydroelectric plants: Furnas, Emborcação, Sobradinho, Agua Vermelha, and Passo Real. The datasets are available from 1931 to 2015, totaling 85 years or 1020 monthly samples. Each sample refers streamflow in m^3/s. These data are public, being available on the website of the National Operator of the Electric System (ONS) [74]. We separated each series on three groups:

- *Training*, from January 1st, 1931 to December 31st, 1995 (780 samples);
- *Validation*, from January 1st, 1996 to December 31st, 2005 (120 samples); and
- *Test*, from January 1st, 2006 to December 31st, 2015 (120 samples).

The mean and standard deviation of all series are available in Table 2. Note their distinct statistical and consequent hydrological behavior, enabling a broader analysis of the results.

Table 2. Mean and standard deviation of each series.

	Complete Series		**Test Set**	
Series	**Mean (m^3/s)**	**S. Deviation (m^3/s)**	**Mean (m^3/s)**	**S. Deviation (m^3/s)**
Furnas	912.1225	613.5036	803.6833	611.6814
Emborcação	480.6578	360.3957	447.7333	355.7428
Sobradinho	2.6062×10^3	1.9412×10^3	1.9607×10^3	1.5001×10^3
Agua Vermelha	2.0773×10^3	1.2957×10^3	1.9635×10^3	1.2668×10^3
Passo Real	208.6216	169.7734	228.0083	167.1326

6.1. Predictors

In this section, we briefly describe the forecasting models used in this work. We consider as predictors two methods: the autoregressive linear model (AR) from the Box and Jenkins methodology [50], and the extreme learning machines neural network (ELM), as the nonlinear representative. Note that when using the approach with 12 predictors, the linear method is called the Periodic Autoregressive model (PAR).

The AR approach linearly weights p past values $\boldsymbol{u}_t = \left[u_{t-1}, u_{t-2}, \ldots, u_{t-p}\right]$ of a time series to provide a future response y_t. Considering that the values of vector $\boldsymbol{u}$ are stationary, (30) explicates such a process:

$$y_t = \varphi_1 u_{t-1} + \varphi_2 u_{t-2} + \ldots + \varphi_p u_{t-p} + a_t \tag{30}$$

where φ_p are the free coefficients of the model.

A significant advantage of this method is the possibility of calculating its coefficients using a close form approach named Yule–Walker equations. This means that, using the same set of inputs, the model

always converges to the same output. This method guarantees the minimum MSE between the output and the desired response.

The standard AR considers just one model to predict all values of the time series. However, it is possible to extend the AR to series, which presents variations in its structure [75], using the periodic autoregressive model (PAR). According to Hippel and McLeod [76], some historical series, such as hydrological ones with seasonal behavior, present an autocorrelation structure linked to the time delay between observations, and the observed period. In this sense, we can address one predictor for each month, the core of the PAR model. For monthly streamflow forecasting, we use 12 predictors, each one adjusted to predict the samples for each month [11].

The second forecasting model addressed is the extreme learning machine (ELM). ELMs are feedforward neural networks like the traditional multilayer perceptron (MLP), with only one intermediate layer. However, the training process differentiates them since the weights of the neurons in the hidden layer are randomly and independently determined. The training process does not adjust the weights of this layer, but only those of the output. The optimal values of the weights are typically calculated analytically since the training involves solving a linear regression problem [77]. Thus, there is no need to calculate derivatives, back-propagate error signals, or use iterative algorithms, which reduces the computational cost involved in the training process.

Bartlett [78] obtained an important theoretical result. The author proved that controlling the norm of synaptic weights is more relevant in terms of the generalization capability of a neural model than controlling the number of neurons in the middle layer. This leads to important evidence that an improvement occurs when the parameter vector has a minimum norm, so the effective number of neurons in the intermediate layer will be defined by the configuration of the weights of the output layer.

Given this statement, ELM presents a guarantee of good generalization effectively given by the weights of the output layer, and the weights of the intermediate layer can be defined at random. Because of this, the network's training becomes linear in relation to the adjustable parameters for supervised training. The generalized Moore–Penrose operator is the most important candidate for solving this problem in the literature [79,80].

In this work, we address the neural network following the same premises of the AR and PAR models: just one ELM for the complete series (annual approach), and 12 ELMs, one adjusted for each month. We highlight that such method was chosen because it presented good results in monthly seasonal streamflow series forecasting, overcoming other neural models [2,12,43–45,80].

6.2. Computational Results

This investigation aims to analyze the quality of the predictions regarding the use of the aforementioned variable selection techniques: filters, wrappers, and bio-inspired metaheuristics. We consider as predictor the autoregressive model (AR), periodic autoregressive model (PAR) [50], and the extreme learning machines neural network (ELM) considering the annual and monthly approaches.

The purpose of these simulations is to find the input selection model that is more suitable for a linear and a non-linear methodologies. Note that the wrappers for AR and PAR models take into account the assessment of the fit of the training set, while the training error of ELMs are not as important because we are interested in the smallest generalization error.

The maximum number of inputs or delays allowed is six, as models of higher orders increase the possibility of negative auto-regressive coefficients [11]. Siqueira et al. [2,12] and Stedinger [52] defend this premise.

The computational results regarding the mean square error (MSE) and mean absolute error (MAE) in the real and deseasonalized (MESd and MAEd) domains for one step ahead are in Tables 3–5. The acronym "Lf" means the linear filter approach based on the partial autocorrelation function (the traditional PACF and the Stedinger's approach PACF-Sted.), "Nf" corresponds to the non-linear filters developed using the mutual information principle (MI, PMI, and N-MRMCR-MI), "WR", the wrapper

method considering as evaluation function the BIC, AIC, and MSE, and "M" the metaheuristics, GA and PSO.

In the AR (Table 3) and PAR (Table 4) cases, we presented the error for training and test sets, while for the single and monthly ELM (Table 5), we just show the errors for the test, because we are interested in analyzing the generalization capability of such response. To the ELM, the results are an average of 30 simulations. The best performances in the test set regarding the MSE in the real domain are highlighted using shades of gray. Additionally, in Appendix A we explicate the inputs selected for all case studies in Tables A1–A5.

Table 3. Results of the variable selection using the AR model.

Variable Selection			Test				Training			
			MSE	MAE	MSEd	MAEd	MSE	MAE	MSEd	MAEd
FURNAS	WR	BIC	109,014	210.77	0.4330	0.5078	**96,033**	**196.58**	**0.4490**	**0.4907**
		AIC	109,014	210.77	0.4330	0.5078	**96,033**	**196.58**	**0.4490**	**0.4907**
		WRAPPER-MSE	107,962	208.81	0.4224	0.4992	97,037	195.55	0.4459	0.4859
	Lf	FACPPe	**107,551**	**209.32**	**0.4259**	**0.5043**	96,646	195.75	0.4464	0.4871
		FACPPe-Sted.	**107,551**	**209.32**	**0.4259**	**0.5043**	96,646	195.75	0.4464	0.4871
	Nf	MI	108,083	209.65	0.4252	0.5015	96,702	195.43	0.4452	0.4856
		PMI	108,083	209.65	0.4252	0.5015	96,702	195.43	0.4452	0.4856
		N-MRMCR-MI	108,083	209.65	0.4252	0.5015	96,702	195.43	0.4452	0.4856
	M	GA	107,962	208.81	0.4224	0.4992	97,037	195.55	0.4459	0.4859
		PSO	107,962	208.81	0.4224	0.4992	97,037	195.55	0.4459	0.4859
EMBORCAÇÃO	WR	BIC	51,745	139.36	0.5487	0.5716	40,456	119.79	0.4838	0.5131
		AIC	51,745	139.36	0.5487	0.5716	40,456	119.79	0.4838	0.5131
		WRAPPER-MSE	51,745	139.36	0.5487	0.5716	40,456	119.79	0.4838	0.5131
	Lf	FACPPe	**50,408**	**138.01**	**0.5353**	**0.5613**	39,953	119.40	0.4790	0.5102
		FACPPe-Sted.	**50,408**	**138.01**	**0.5353**	**0.5613**	39,953	119.40	0.4790	0.5102
	Nf	MI	50,559	138.24	0.5397	0.5602	**39,880**	**119.20**	**0.4768**	**0.5081**
		PMI	50,559	138.24	0.5397	0.5602	**39,880**	**119.20**	**0.4768**	**0.5081**
		N-MRMCR-MI	50,559	138.24	0.5397	0.5602	**39,880**	**119.20**	**0.4768**	**0.5081**
	M	GA	51,745	139.36	0.5487	0.5716	40,456	119.79	0.4838	0.5131
		PSO	51,745	139.36	0.5487	0.5716	40,456	119.79	0.4838	0.5131
SOBRADINHO	WR	BIC	836,738	568.04	0.3071	0.4408	1,032,181	580.83	0.3895	0.4366
		AIC	836,738	568.04	0.3071	0.4408	1,032,181	580.83	0.3895	0.4366
		WRAPPER-MSE	836,738	568.04	0.3071	0.4408	1,032,181	580.83	0.3895	0.4366
	Lf	FACPPe	836,738	568.04	0.3071	0.4408	1,032,181	580.83	0.3895	0.4366
		FACPPe-Sted.	863,796	577.30	0.3196	0.4515	1,020,492	578.13	0.3910	0.4403
	Nf	MI	**828,142**	**566.45**	**0.3043**	**0.4375**	**994,408**	**573.77**	**0.3837**	**0.4350**
		PMI	**828,142**	**566.45**	**0.3043**	**0.4375**	**994,408**	**573.77**	**0.3837**	**0.4350**
		N-MRMCR-MI	**828,142**	**566.45**	**0.3043**	**0.4375**	**994,408**	**573.77**	**0.3837**	**0.4350**
	M	GA	836,738	568.04	0.3071	0.4408	1,032,181	580.83	0.3895	0.4366
		PSO	836,738	568.04	0.3071	0.4408	1,032,181	580.83	0.3895	0.4366
AGUA VERMELHA	WR	BIC	417,720	404.35	0.4097	0.4826	378,866	394.30	0.4095	0.4780
		AIC	417,720	404.35	0.4097	0.4826	378,866	394.30	0.4095	0.4780
		WRAPPER-MSE	417,720	404.35	0.4097	0.4826	378,866	394.30	0.4095	0.4780
	Lf	FACPPe	**409,613**	**401.00**	**0.4052**	**0.4803**	379,329	392.98	0.4078	0.4752
		FACPPe-Sted.	**409,613**	**401.00**	**0.4052**	**0.4803**	379,329	392.98	0.4078	0.4752
	Nf	MI	415,465	404.50	0.4062	0.4828	378,197	392.12	0.4065	0.4749
		PMI	413,991	403.40	0.4119	0.4886	**369,658**	**393.50**	**0.4149**	**0.4819**
		N-MRMCR-MI	413,991	403.40	0.4119	0.4886	378,197	392.12	0.4065	0.4749
	M	GA	415,465	404.50	0.4062	0.4828	378,197	392.12	0.4065	0.4749
		PSO	415,465	404.50	0.4062	0.4828	378,197	392.12	0.4065	0.4749
PASSO REAL	WR	BIC	14,996	88.65	0.6570	0.5969	16,637	86.70	0.6490	0.5718
		AIC	14,996	88.65	0.6570	0.5969	16,637	86.70	0.6490	0.5718
		WRAPPER-MSE	14,996	88.65	0.6570	0.5969	16,637	86.70	0.6490	0.5718
	Lf	FACPPe	**14,523**	**87.74**	**0.6397**	**0.5914**	16,497	86.32	0.6415	0.5696
		FACPPe-Sted.	**14,523**	**87.74**	**0.6397**	**0.5914**	16,497	86.32	0.6415	0.5696
	Nf	MI	14,632	88.16	0.6447	0.5956	**16,478**	**86.08**	**0.6398**	**0.5676**
		PMI	14,632	88.16	0.6447	0.5956	**16,478**	**86.08**	**0.6398**	**0.5676**
		N-MRMCR-MI	14,632	88.16	0.6447	0.5956	**16,478**	**86.08**	**0.6398**	**0.5676**
	M	GA	14,996	88.65	0.6570	0.5969	16,637	86.70	0.6490	0.5718
		PSO	14,996	88.65	0.6570	0.5969	16,637	86.70	0.6490	0.5718

We also applied the Friedman's test to evaluate if the results are significantly distinct [57]. As expected, for the cases that the same set of inputs are selected, there is no statistical difference between the results for the same model. In almost all the other cases, the *p*-values achieved were close to zero, indicating that change the inputs leads to different conclusions. We discuss the exceptions below.

Table 4. Results of the variable selection using PAR model.

Variable Selection			Test				Training			
			MSE	MAE	MSEd	MAEd	MSE	MAE	MSEd	MAEd
FURNAS	WR	BIC	117,347	207.37	0.4096	0.4819	94,879	190.77	0.4124	0.4725
		AIC	120,629	211.50	0.4343	0.4957	93,403	189.09	0.4041	0.4677
		WRAPPER-MSE	128,415	215.49	0.4546	0.5046	87,842	183.45	0.3846	0.4556
	Lf	PACF	118,744	211.52	0.4113	0.4882	102,317	198.26	0.4499	0.4899
		PACF-Sted.	**117,144**	**206.70**	**0.4055**	**0.4788**	94,824	190.41	0.4112	0.4709
	Nf	MI	120,121	211.57	0.4302	0.4963	92,821	187.35	0.3973	0.4620
		PMI	122,682	215.72	0.4442	0.5055	96,930	193.12	0.4502	0.4822
		N-MRMCR-MI	135,260	224.64	0.4811	0.5271	95,500	187.39	0.3999	0.4619
	M	GA	128,680	216.09	0.4601	0.5073	**87,837**	**183.33**	**0.3845**	**0.4550**
		PSO	133,171	228.39	0.4794	0.5279	109,291	201.15	0.4856	0.5007
EMBORCAÇÃO	WR	BIC	46,356	128.98	0.5034	0.5350	35,421	112.66	0.4377	0.4817
		AIC	46,356	129.00	0.5034	0.5352	35,415	112.57	0.4370	0.4806
		WRAPPER-MSE	51,529	134.64	0.5460	0.5489	**33,362**	**109.12**	**0.4188**	**0.4702**
	Lf	PACF	50,251	134.96	0.5526	0.5612	42,622	123.54	0.5596	0.5418
		PACF-Sted.	46,195	129.23	0.5093	0.5444	35,710	113.81	0.4535	0.4914
	Nf	MI	47,918	130.25	0.5094	0.5353	39,096	115.99	0.4588	0.4851
		PMI	48,392	129.72	0.5183	0.5361	40,967	118.63	0.4913	0.5010
		N-MRMCR-MI	**46,088**	**128.22**	**0.4982**	**0.5343**	35,593	113.05	0.4419	0.4836
	M	GA	51,529	134.64	0.5460	0.5489	**33,362**	**109.12**	**0.4188**	**0.4702**
		PSO	58,768	153.16	0.9161	0.6795	48,079	136.61	0.8187	0.6266
SOBRADINHO	WR	BIC	650,374	507.00	0.2791	0.4097	850,140	530.47	0.3560	0.4133
		AIC	642,631	496.61	0.2706	0.3972	847,763	528.97	0.3530	0.4118
		WRAPPER-MSE	675,439	510.13	0.3071	0.4194	829,497	523.19	0.3450	0.4061
	Lf	PACF	724,099	546.38	0.3361	0.4462	1,021,024	577.88	0.4665	0.4523
		PACF-Sted.	666,690	513.82	0.2896	0.4175	886,340	540.87	0.3671	0.4207
	Nf	MI	**628,672**	**495.94**	**0.2923**	**0.4124**	847,892	530.37	0.3519	0.4104
		PMI	665,958	511.55	0.2948	0.4171	886,511	544.87	0.3852	0.4340
		N-MRMCR-MI	682,074	513.84	0.3034	0.4180	**824,000**	**519.21**	**0.3399**	**0.4010**
	M	GA	675,494	510.64	0.3074	0.4207	829,499	523.20	0.3450	0.4061
		PSO	755,372	564.36	0.3402	0.4509	1,181,095	632.30	0.5290	0.4889
AGUA VERMELHA	WR	BIC	438,074	401.41	0.4075	0.4729	357,604	380.46	0.3774	0.4614
		AIC	439,586	402.88	0.4161	0.4769	356,951	379.09	0.3742	0.4587
		WRAPPER-MSE	473,968	409.23	0.4437	0.4881	335,448	366.74	0.3564	0.4461
	Lf	PACF	469,836	407.75	0.4179	0.4732	404,467	393.70	0.4056	0.4728
		PACF-Sted.	**432,799**	**393.57**	**0.3960**	**0.4611**	359,174	382.19	0.3817	0.4645
	Nf	MI	459,970	408.37	0.4300	0.4836	379,638	380.90	0.3817	0.4566
		PMI	443,387	399.30	0.4188	0.4711	360,565	384.67	0.3838	0.4673
		N-MRMCR-MI	436,274	395.68	0.4029	0.4664	357,410	380.80	0.3781	0.4618
	M	GA	476,365	409.64	0.4450	0.4884	**335,047**	**366.40**	**0.3561**	**0.4459**
		PSO	775,418	560.05	0.7722	0.6972	627,743	500.78	0.7843	0.6385
PASSO REAL	WR	BIC	16,793	93.49	0.7945	0.6449	15,864	85.70	0.6228	0.5684
		AIC	15,601	88.50	0.7664	0.6198	15,299	84.28	0.6022	0.5590
		WRAPPER-MSE	15,584	91.70	0.7454	0.6395	14,770	82.82	0.5797	0.5474
	Lf	PACF	15,924	90.39	0.8031	0.6296	15,012	84.05	0.6062	0.5605
		PACF-Sted.	15,522	89.04	0.7557	0.6151	14,976	83.78	0.6018	0.5576
	Nf	MI	**14,982**	**90.73**	**0.7152**	**0.6300**	16,107	85.61	0.6336	0.5671
		PMI	15,282	87.41	0.7423	0.6038	16,013	85.63	0.6344	0.564
		N-MRMCR-MI	15,474	90.55	0.7402	0.6306	**14,638**	**82.25**	**0.5724**	**0.5428**
	M	GA	15,779	92.12	0.7551	0.6424	14,769	82.99	0.5796	0.5486
		PSO	17,895	101.47	0.8774	0.7130	22,018	103.18	0.8769	0.6816

Table 5. Results of the variable selection using ELM.

Variable Selection			Monthly Approach				Annual Approach			
			MSE	MAE	MSEd	MAEd	MSE	MAE	MSEd	MAEd
FURNAS	WR	BIC	124,391	212.47	0.4320	0.4943	***123,754***	***220.19***	***0.4597***	***0.5307***
		AIC	123,550	210.02	0.4264	0.4878	126,426	219.67	0.4503	0.5195
		WRAPPER-MSE	**119,067**	**206.45**	**0.4060**	**0.4809**	***123,745***	***217.80***	***0.4455***	***0.5171***
	Lf	PACF	126,594	222.32	0.4623	0.5236	129,323	219.72	0.4451	0.5138
		PACF-Sted.	121,806	212.11	0.4433	0.4997	126,768	216.79	0.4509	0.5110
	Nf	MI	130,637	230.24	0.5115	0.5656	***124,105***	***220.33***	***0.4665***	***0.5279***
		PMI	126,445	217.60	0.4769	0.5207	***126,349***	***221.97***	***0.4611***	***0.5287***
		N-MRMCR-MI	140,451	238.15	0.5424	0.5842	126,507	222.67	0.4650	0.5330
	M	GA	137,304	230.58	0.5063	0.5448	135,978	232.35	0.4963	0.5594
		PSO	132,599	228.87	0.4829	0.5377	131,426	226.52	0.4823	0.5435
EMBORCAÇÃO	WR	BIC	38,143	114.51	0.4495	0.4956	48,513	130.03	0.5333	0.5586
		AIC	41,110	116.21	0.4657	0.4944	***45,459***	***127.26***	***0.5158***	***0.5501***
		WRAPPER-MSE	**37,551**	**118.56**	**0.4335**	**0.5001**	***44,936***	***129.15***	***0.5137***	***0.5557***
	Lf	PACF	44,227	124.36	0.5122	0.5355	***45,994***	***130.15***	***0.5153***	***0.5563***
		PACF-Sted.	48,543	130.42	0.5395	0.5526	***44,690***	***129.01***	***0.5095***	***0.5556***
	Nf	MI	49,315	131.44	0.5641	0.5662	***45,169***	***130.89***	***0.5126***	***0.5565***
		PMI	52,571	132.20	0.5707	0.5564	***44,094***	***128.54***	***0.5061***	***0.5511***
		N-MRMCR-MI	47,931	129.26	0.5398	0.5459	***45,707***	***129.50***	***0.5169***	***0.5558***
	M	GA	54,659	142.88	0.6225	0.6017	***45,434***	***128.47***	***0.5166***	***0.5515***
		PSO	50,211	130.82	0.5667	0.5644	***45,298***	***130.20***	***0.5170***	***0.5595***
SOBRADINHO	WR	BIC	642,185	519.69	0.2954	0.4329	669,441	534.37	0.3209	0.4632
		AIC	590,254	492.20	0.2979	0.4298	**657,405**	**530.22**	**0.3166**	**0.4532**
		WRAPPER-MSE	**587,680**	**495.73**	**0.2945**	**0.4250**	672,783	531.90	0.3229	0.4567
	Lf	PACF	696,480	530.92	0.3376	0.4506	718,719	549.40	0.3366	0.4676
		PACF-Sted.	747,932	550.63	0.3561	0.4591	690,307	549.76	0.3510	0.4872
	Nf	MI	692,277	533.35	0.3526	0.4587	668,656	531.83	0.3187	0.4527
		PMI	746,916	546.82	0.3647	0.4621	694,596	540.00	0.3234	0.4549
		N-MRMCR-MI	828,437	582.14	0.4819	0.3842	710,649	542.32	0.3316	0.4583
	M	GA	728,468	563.81	0.3400	0.4674	698,381	555.96	0.3501	0.4869
		PSO	773,147	571.99	0.3657	0.4738	712,829	558.87	0.3655	0.4920
AGUA VERMELHA	WR	BIC	408,982	384.66	0.3646	0.4528	443,959	394.45	0.4055	0.4727
		AIC	411,485	387.40	0.3701	0.4583	**436,790**	**393.10**	**0.3946**	**0.4661**
		WRAPPER-MSE	**374,264**	**375.43**	**0.3412**	**0.4429**	436,981	394.69	0.3952	0.4664
	Lf	PACF	436,494	412.25	0.4109	0.4864	439,692	401.17	0.4025	0.4759
		PACF-Sted.	419,472	401.62	0.4040	0.4813	458,381	406.68	0.4131	0.4813
	Nf	MI	423,961	426.85	0.4565	0.5274	453,903	420.36	0.4288	0.5065
		PMI	434,519	410.82	0.4420	0.5026	432,417	394.64	0.3955	0.4687
		N-MRMCR-MI	417,154	397.18	0.4020	0.4787	458,725	422.00	0.4344	0.5076
	M	GA	502,689	437.62	0.4673	0.5201	449,919	405.97	0.4142	0.4874
		PSO	478,617	440.21	0.4526	0.5271	439,377	402.75	0.4022	0.4806
PASSO REAL	WR	BIC	12,768	79.70	0.6382	0.5549	15,859	89.78	0.7367	0.6079
		AIC	12,964	78.81	0.6592	0.5490	15,866	89.78	0.7366	0.6078
		WRAPPER-MSE	**11,828**	**78.42**	**0.6033**	**0.5488**	15,435	87.80	0.7277	0.5962
	Lf	PACF	16,288	91.41	0.7772	0.6218	**15,257**	**86.40**	**0.7278**	**0.5884**
		PACF-Sted.	16,435	89.65	0.7850	0.6116	15,351	86.55	0.7320	0.5893
	Nf	MI	15,059	86.49	0.7196	0.5944	16,074	89.96	0.7612	0.6170
		PMI	15,400	87.35	0.7517	0.5979	16,146	91.00	0.7584	0.6207
		N-MRMCR-MI	16,635	91.48	0.7723	0.6274	16,208	90.10	0.7683	0.6168
	M	GA	17,035	90.31	0.8134	0.6140	16,258	91.15	0.7626	0.6206
		PSO	15,589	86.86	0.7344	0.5964	16,270	91.04	0.7731	0.6217

The critical analysis regarding the results achieved by the AR model reveals interesting behaviors (Table 3). For Furnas and Emborcação time series, there was no perfect correspondence between the best performance regarding the MSE and MAE in the real and deseasonalized domains' errors. In such cases, we assumed the best predictor with the smallest MSE in the real space, following the premises already stated in previous works [2,12,80].

The general analysis showed other relevant issues: for training or test, at least two variable methods led to the same performances since they selected the same subset of inputs, except for the Agua Vermelha training set. As the AR optimized by the Yule-Walker equations presents a closed-form solution, the same input vector necessarily leads to the same output responses. This behavior can occur since just one model is adjusted, and the number of inputs is limited to six.

Likewise, these draws happened for the same class of variable selection method. Note, for example, for Emborcação, Sobradinho, and Passo Real, the smallest training error in the MSE sense was related to the non-linear filters, while for Furnas were the wrappers based on BIC and AIC.

However, we observed an intriguing behavior: the best performances were related to distinct variable selection methods for training and test sets (Table 3). Following some literature regarding monthly seasonal streamflow series forecasting [2,12], one should state the best variable selection method related to the error in the test set. However, the analysis of the training set presented the search capability of the methods. Indeed, the ideal behavior would be the same VS approach for both. The training set was better adjusted for MI filters in 4 cases, and the wrapper (BIC and AIC) in one. PACF best fitted the test sets four of five times, and by the MI filters, once.

Analyzing the inputs selected for AR model in Tables A1–A5 in Appendix A, one can note that the models that fitted better in training set selected six lags for Emborcação, Sobradinho, and Passo Real. For Furnas and Agua Vermelha, two entries. Considering the test set, the PACF approaches, in general, selected three or four inputs. As expected, the MI methods usually chose more lags than the PACF, since they detect non-linear relations. As the linear approach achieved four of five best results, we can affirm that include all inputs in the AR model may tend to a configuration with less generalization capability.

Although the bio-inspired metaheuristics present an elevated search capability, for the AR neither, PSO nor GA achieved some of the best performances (Table 3).

Unlike the AR case, the PAR model's results presented a draw just for the training set of Emborcação (Table 4). It happened because both GA and wrapper-MSE found the same set of inputs (see Table A2). Considering the optimization of 12 models simultaneously, totaling up to 72 free parameters, it is more likely that the variable selection models achieve distinct configurations.

To the PAR model (Table 4), the best VS method presented homogeneity regarding the four error metrics. On the other hand, we noted again that the best performance for training did not present correspondence with the test set. In the training set, the GA stood out, achieving the smallest errors for Emborcação, Furnas, and together with wrapper-MSE for Agua Vermelha. For the other scenarios, we highlight the N-MRMCR-MI. In the test set, just filters reached the smallest errors: PACF-Sted. (Furnas and Agua Vermelha), N-MRMCR-MI (Emborcação), and MI for the others.

Tables A1–A5 reveal the GA and wrapper-MSE often selected five or six inputs for all months. In the comparison of wrapper methodologies, a pattern could be noticed, since the BIC selected fewer inputs than AIC. In practice, we observed the influence of each type of penalty regarding the insertion of new entries.

The results achieved by the ELM considering the annual approach, summarized in Table 5, must be discussed considering not just the numerical values of the errors, but also the statistical difference between them. Additionally, it is important to highlight that when applying neural networks, there is no interest in evaluating the training error, because we are looking for the configuration to achieve the highest generalization capability. Therefore, we discuss just the error in the test set.

In the annual models, we once again noted some ties for Furnas and Emborcação. In addition to Table 5 presenting distinct numerical values, the obtained p-value for the Friedman test was higher than 0.05, as expected, since the set of selected inputs was the same (see Tables A1–A5). As the initialization of the weights of an ELM is random, the outputs are distinct, but the results are close. It is the reason we must run the algorithm at least 30 times.

For Furnas, wrapper (AIC and MSE), and the non-linear filters (MI and PMI) presented the best results, selecting the same entries. For Emborcação, a further discussion must be done. Except for the

wrapper-BIC, all methods led to the same performances statistically, according to the Friedman test. However, one can find four distinct sets of inputs in Table A2. It can happen since the approximation capability of a neural network is elevated, and these models are universal approximators. Regarding the inputs, note that lags 1 and 4 belong to all subsets.

For the other series, the AIC stood out for Agua Vermelha and Sobradinho, while the PACF and PACF-Sted., for Passo Real (same set of inputs). The AIC method selected two inputs for all cases.

As observed for the linear prediction models, the MI methods tended to select more lags than the linear filters (see Tables A1–A5). The PMI presented fewer inputs than their non-linear counterparts. In general, it seems clear that inserting too many inputs does not necessarily lead to an increase in performance, especially in the test set.

The monthly prediction using ELM, similar to the linear case, showed error values with a standard pattern. For all scenarios, the wrapper-MSE found the best performances, although the smallest MAE for Sobradinho is related to AIC. In a few cases, the best error metrics converge to the same predictor. Except for the tie in Emborcação, the metaheuristics did not achieve the best errors.

6.3. Discussion

Table 6 displays how many times each VS method led to the best performance, according to the models depicted in Tables 3–5: AR (training and test), PAR (training and test), annual ELM (test), and monthly ELM (test). Due to the statistical similarity between some of the best results, provided by Friedman test, we indicated the number of VS methods that achieved similar performance regarding each predictor between parentheses. For example, for the AR in the training set, the PMI reached the best result four times, once alone and three times together with two more VS methods. In bold and underlined it is highlighted the number of times some approach was the single best.

Table 6. Number of best results by approach.

	Models					
VS Method	**AR Train**	**AR Test**	**PAR Train**	**PAR Test**	**ELM Annual**	**ELM Monthly**
BIC	1(+1)	-	-	-	1(+3)	-
AIC	1(+1)	-	-	-	**2**; 1(+8)	-
WRAPPER-MSE	-	-	1(+1)	-	1(+3); 1(+8)	**5**
PACF	-	1(+1); 1(+1); 1(+1); 1(+1)	-	-	**1**; 1(+8)	-
PACF-Sted.	**-**	1(+1); 1(+1); 1(+1); 1(+1)	**-**	**2**	1(+8)	**-**
MI	1(+2); 1(+2); 1(+2)	1(+2)	-	**2**	1(+3); 1(+8)	-
PMI	**1**; 1(+2); 1(+2); 1(+2)	1(+2)	-	**-**	1(+3); 1(+8)	-
N-MRMCR-MI	1(+2); 1(+2); 1(+2)	1(+2)	**1**	**1**	1(+8)	-
GA	-	-	**3**; 1(+1)	-	1(+8)	-
PSO	-	-	-	-	1(+8)	-

The general analysis considering all results allows some interesting observations. In annual approaches, we noted many draws, unlike in the monthly approach. It is plausible since only one predictor is adjusted. The monthly models presented more sprayed performances due to the number of models fitted for each series (see Tables 3–5).

In most cases, the PACF and PACF - Sted. presented the same set of inputs (see Appendix A, Tables A1–A5). It indicates that, for hydrologic series, the hypothesis of dependency of consecutive delays makes sense. These methods were highlighted to the test set of the AR and PAR models (Table 6).

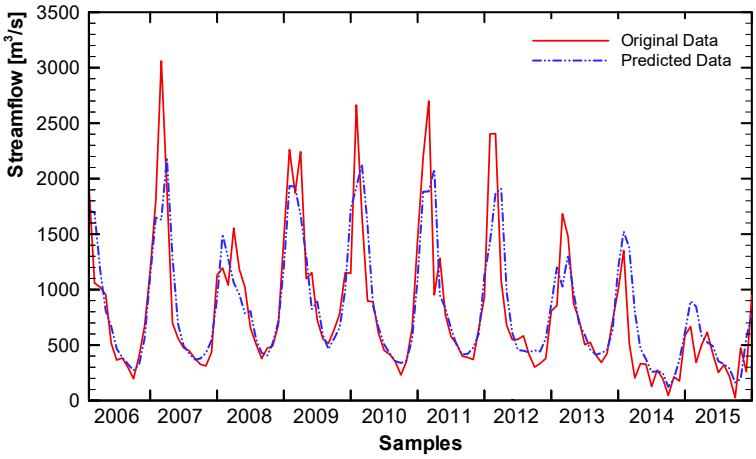
Considering the non-linear filters for annual models, these approaches presented as inputs all the six lags, except for Agua Vermelha. In general, comparing the MI-methods to the monthly cases, the PMI tended to select fewer inputs, while the N-MRMCR-MI selected more inputs, except for Agua

Vermelha (see Tables A1–A5). The MI-based approaches are avid for data since the estimation of the probability density function (PDF) gets better; the more data are available. That is why for annual models, such an approach tends to behave better than the monthly models. The AR was better adjusted for the training set using MI. Additionally, the N-MRMCR-MI won alone twice (Table 6).

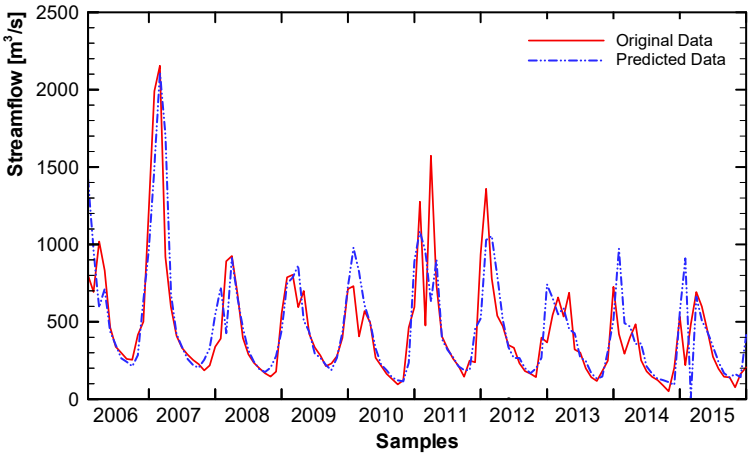
Regarding the wrapper methodology, as expected, the MSE-based selected more inputs then BIC and AIC, since there is no penalty in introducing new lags. The BIC selected fewer entries than the AIC, due to its strong penalty function (see Tables A1–A5). However, The BIC did not overcome the others by itself, unlike the AIC, which was the winner twice for ELM annual model. In summary, the wrapper approach stood out for the neural networks (Table 6).

Regarding the metaheuristics, just for the PAR in the training set, the GA achieved most of the best performances. These are powerful methods to deal with binary optimization problems like variable selection, and we believe this approach could be competitive [68,69]. Additionally, the computational cost was higher than the wrappers (Table 6). Moreover, we can also notice that for four of five series, the monthly ELM achieved the best general performances (Table 3, Table 4 and Table 5). It is particularly relevant since, in current days, one can still find the massive use of linear models in the literature [2,12].

Finally, we state that variable selection is a complex problem, little explored in the context of monthly seasonal streamflow forecasting. The variety among the responses proves the task's inherent difficulties, which present elevated economic and environmental impacts. As can be seen, the models' efficiency is greatly influenced by lags choice. In Figure 8, we present the best predictions achieved by the winner in each scenario: AR for Furnas and monthly ELM for the other cases.

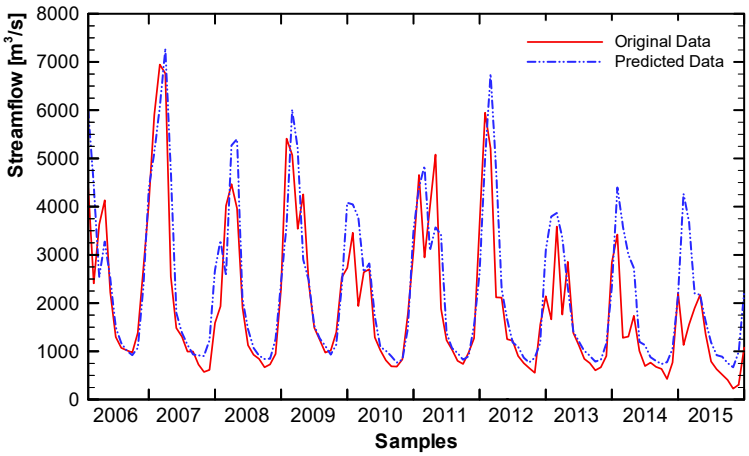


(**a**) Furnas

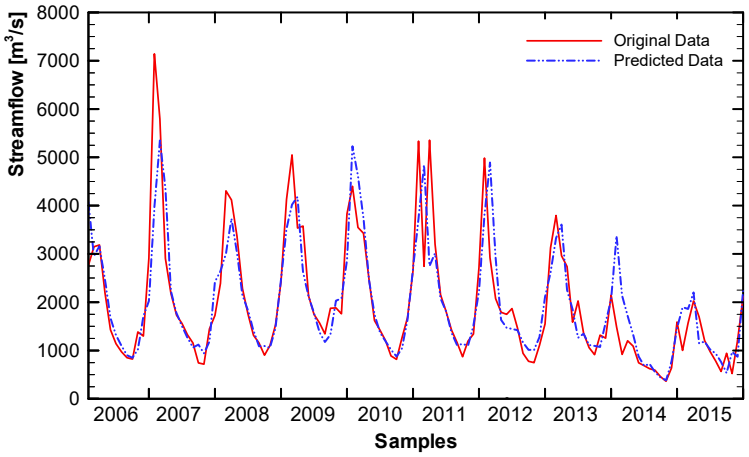


(**b**) Emborcação

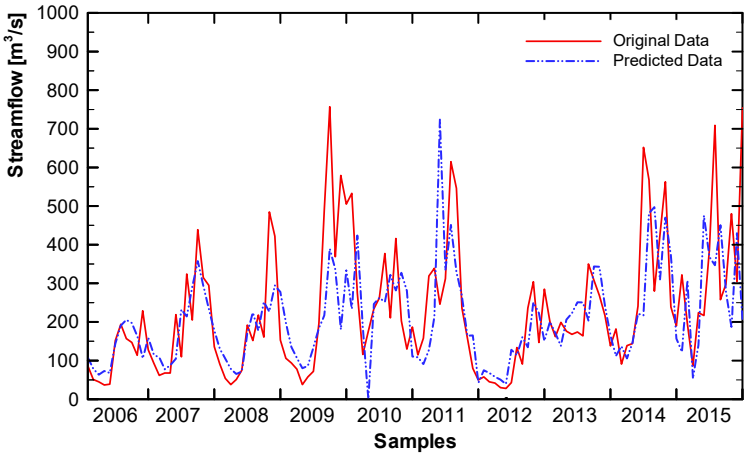
Figure 8. *Cont.*



(**c**) Sobradinho



(**d**) Agua Vermelha



(**e**) Passo Real

Figure 8. Best prediction by plant for $p = 1$.

7. Conclusions

This work performed an extensive investigation on the variable selection methods to determine the best subsets of inputs to increase the accuracy in the monthly seasonal streamflow series forecasting tasks. Most of the specialized literature focus on finding the best predictor, but the selection of the inputs presents an essential step for forecasting.

We addressed wrappers, linear and non-linear filters, and bio-inspired metaheuristics. The wrapper methodology can evaluate the quality of this subset under several criteria, including:

- Mean square error (MSE);
- Bayesian information criterion (BIC);
- Bayesian information criterion (AIC).
- The linear filters used were the:
- Partial autocorrelation function (PACF);
- PACF using the Stedinger [52] approach for hydrological series.
- The nonlinear filters addressed were:
- Mutual Information (MI);
- Partial mutual information (PMI); and
- Normalization of maximum relevance and minimum common redundancy mutual information (N-MRMCR-MI).

Regarding the metaheuristics, we used binary versions of the:

- Particle swarm optimization (PSO); and
- Genetic algorithm (GA).

We performed computational tests with the predictions made for five monthly series related to hydroelectric plants using the autoregressive linear model (AR), and extreme learning machine neural network (ELM) as predictors. We also addressed two forecasting approaches, the use of only one predictor for the whole series, and the use of 12 predictors, each one adjusted for each month.

The main findings of this investigation are:

- The selected lags were very diverse depending on the method, especially for the monthly case;
- For the annual approaches, some draws could be found;
- The linear models perform better with filters;
- The wrapper is the best choice for the neural network; and
- Regarding the forecasting methods, the monthly ELM achieved the best error values.

These findings are especially important for countries where the power generation is mainly from hydroelectric plants, since this is the most important renewable power source in the world. Additionally, such investigation can contribute to energetic planning, water availability, and pricing strategies for the power productive chains.

Future works can be developed considering these approaches for other series of renewable inputs for power generation, like wind power [36,38]. Other problems related to energy generation as an estimation of methane production and biogas efficiency [81] could be treated and simulated using a similar methodology. Furthermore, other metaheuristics can be addressed, since there is a vast repertoire of possibilities being developed in the last years, such as differential evolution and artificial bee colony, among others.

Author Contributions: Conceptualization: H.S., S.L.S.J., Y.d.S.T., T.A.A., M.H.N.M., and I.L.; methodology: H.S., S.L.S.J., Y.d.S.T., and T.A.A.; software: H.S., I.L., D.S.O.J., and M.H.N.M.; validation: H.S., P.S.G.d.M.N., J.F.L.d.O., and M.H.N.M.; formal analysis: P.S.G.d.M.N., M.H.N.M., and J.F.L.d.O.; investigation: Y.d.S.T., and T.A.A.; resources: M.H.N.M; writing—original draft preparation: H.S., Y.d.S.T., M.H.N.M., I.L., L.A.S., and A.C.; writing—review and editing: M.d.A.L.F., L.A.S., T.A.A., and A.C.; illustrations: T.A.A.; visualization: P.S.G.d.M.N., T.A.A., and J.F.L.d.O.; supervision: L.A.S., A.C., and M.H.N.M; project administration: M.H.N.M; funding acquisition: H.S. and M.H.N.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work received funding and technical support from AES and Associated Companies (of the CPFL, Brookfield, and Global group) as part of the ANEEL PD-0610-1004/2015 project, "IRIS - Integration of intermittent renewables: A simulation model of the operation of the Brazilian electrical system to support planning, operation, commercialization, and regulation", which is part of an R and D program regulated by ANEEL, Brazil. The authors also thank the Advanced Institute of Technology and Innovation (IATI) for its support, the National Institute of Meteorology (INMET) for providing the data, and the Coordination for the Improvement of Higher Education Personnel - Brazil (CAPES)—Financing Code 001, for support.

Acknowledgments: This work was partially supported by the Brazilian agencies CAPES and FACEPE. The authors also thank the Brazilian National Council for Scientific and Technological Development (CNPq), process number 40558/2018-5, and Araucaria Foundation, process number #51497, for their financial support.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

In this appendix we present in Tables A1–A5 the inputs selected by each VS methodology, regarding the five monthtly seasonal streamflow series from five Brazilian hydroelectric plants. Note that the best performance related to Tables 3–5 in the test set are highlighted in bold. For the training set, the best results are in italics and bold.

Table A1. Furnas.

Month		BIC	AIC	W-MSE	PACF	PACF -Sted.	MI	PMI	N-MRMCR -MI	GA	PSO
PAR	**J**	1(1)	1(1)	5(1,3,4,5,6)	**1(1)**	1(1)	1(1)	1(1)	3(1,5,6)	***5(1,3,4,5,6)***	5(12,3,4,6)
	F	1(1)	1(1)	6(1,2,3,4,5,6)	**2(1,6)**	1(1)	2(1,2)	1(1)	6(1,2,3,4,5,6)	***6(1,2,3,4,5,6)***	3(2,4,6)
	M	1(1)	2(1,2)	6(1,2,3,4,5,6)	**2(1,6)**	1(1)	2(1,2)	1(1)	5(1,2,3,4,5,5)	***6(1,2,3,4,5,6)***	2(1,5)
	A	2(1,2)	3(1,2,3)	6(1,2,3,4,5,6)	**2(1,2)**	2(1,2)	4(1,2,3,4)	6(1,2,3,4,5,6)	6(1,2,3,4,5,6)	***6(1,2,3,4,5,6)***	3(1,3,5)
	M	4(1,2,3,4)	4(1,2,3,4)	6(1,2,3,4,5,6)	**3(1,2,3)**	3(1,2,3)	5(1,2,3,4,5)	2(1,3)	4(1,3,4,6)	***6(1,2,3,4,5,6)***	3(1,2,3)
	J	2(1,2)	2(1,2)	6(1,2,3,4,5,6)	**2(1,2)**	2(1,2)	6(1,2,3,4,5,6)	3(1,2,5)	6(1,2,3,4,5,6)	***6(1,2,3,4,5,6)***	2(1,4)
	J	2(1,2)	2(1,2)	6(1,2,3,4,5,6)	**2(1,2)**	2(1,2)	6(1,2,3,4,5,6)	4(1,2,4,5)	6(1,2,3,4,5,6)	***6(1,2,3,4,5,6)***	5(1,2,4,5,6)
	A	1(1)	4(1,2,3,4)	6(1,2,3,4,5,6)	**1(1)**	1(1)	6(1,2,3,4,5,6)	6(1,2,3,4,5,6)	6(1,2,3,4,5,6)	***6(1,2,3,4,5,6)***	4(1,3,5,6)
	S	4(1,2,3,4)	4(1,2,3,4)	6(1,2,3,4,5,6)	**4(1,2,3,4)**	4(1,2,3,4)	6(1,2,3,4,5,6)	1(1)	6(1,2,3,4,5,6)	***6(1,2,3,4,5,6)***	4(1,2,3,6)
	O	2(3,4)	3(3,4,6)	6(1,2,3,4,5,6)	**4(1,2,3,4)**	4(1,2,3,4)	6(1,2,3,4,5,6)	1(2)	6(1,2,3,4,5,6)	***5(1,2,3,4,6)***	4(1,3,4,6)
	N	1(1)	2(1,2)	6(1,2,3,4,5,6)	**2(1,5)**	1(1)	6(1,2,3,4,5,6)	2(1,2)	6(1,2,3,4,5,6)	***5(1,2,3,4,5)***	4(1,2,4,5)
	D	2(1,2)	2(1,2)	6(1,2,3,4,5,6)	**3(1,2,6)**	2(1,2)	4(1,2,3,4)	1(1)	6(1,2,3,4,5,6)	***6(1,2,3,4,5,6)***	2(4,5)
ELM	**J**	1(2)	1(2)	**1(2)**	1(1)	1(1)	1(1)	1(1)	3(1,5,6)	2(1,4)	1(1)
	F	1(1)	1(1)	**1(1)**	2(1,6)	1(1)	2(1,2)	1(1)	6(1,2,3,4,5,6)	2(1,4)	5(1,3,4,5,6)
	M	1(1)	1(1)	**2(3,6)**	2(1,6)	1(1)	2(1,2)	1(1)	5(1,2,3,4,5,5)	5(1,2,3,4,5)	3(1,3,4)
	A	1(5)	1(5)	**1(5)**	2(1,2)	2(1,2)	4(1,2,3,4)	6(1,2,3,4,5,6)	6(1,2,3,4,5,6)	2(1,2)	4(1,2,4,5)
	M	1(2)	1(2)	**1(2)**	3(1,2,3)	3(1,2,3)	5(1,2,3,4,5)	2(1,3)	4(1,3,4,6)	2(2,3)	3(1,2,3)
	J	1(1)	1(1)	**1(1)**	2(1,2)	2(1,2)	6(1,2,3,4,5,6)	3(1,2,5)	6(1,2,3,4,5,6)	1(1)	1(1)
	J	1(1)	1(1)	**1(1)**	2(1,2)	2(1,2)	6(1,2,3,4,5,6)	4(1,2,4,5)	6(1,2,3,4,5,6)	2(1,6)	1(1)
	A	1(1)	1(1)	**1(1)**	1(1)	1(1)	6(1,2,3,4,5,6)	6(1,2,3,4,5,6)	6(1,2,3,4,5,6)	4(1,2,3,5)	1(3)
	S	1(1)	1(1)	**1(1)**	4(1,2,3,4)	4(1,2,3,4)	6(1,2,3,4,5,6)	1(1)	6(1,2,3,4,5,6)	1(1)	2(1,3)
	O	1(1)	1(1)	**1(1)**	4(1,2,3,4)	4(1,2,3,4)	6(1,2,3,4,5,6)	1(2)	6(1,2,3,4,5,6)	1(1)	3(1,2,6)
	N	1(5)	1(5)	**2(2,5)**	2(1,5)	1(1)	6(1,2,3,4,5,6)	2(1,2)	6(1,2,3,4,5,6)	1(1)	3(2,3,5)
	D	1(2)	1(2)	**1(2)**	3(1,2,6)	2(1,2)	4(1,2,3,4)	1(1)	6(1,2,3,4,5,6)	1(2)	2(2,6)
AR		***2(1,2)***	***2(1,2)***	4(1,2,3,5)	**3(1,2,3)**	**3(1,2,3)**	6(1,2,3,4,5,6)	6(1,2,3,4,5,6)	6(1,2,3,4,5,6)	4(1,2,3,5)	4(1,2,3,5)
ELM		1(1)	2(1,4)	**6(1,2,3,4,5,6)**	3(1,2,3)	3(1,2,3)	**6(1,2,3,4,5,6)**	**6(1,2,3,4,5,6)**	**6(1,2,3,4,5,6)**	4(1,2,3,5)	5(1,2,4,5,6)

Table A2. Emborcação.

Month		BIC	AIC	W-MSE	PACF	PACF -Sted.	MI	PMI	N-MRMCR -MI	GA	PSO
PAR	**J**	1(1)	1(1)	***6(1,2,3,4,5,6)***	2(1,6)	1(1)	1(6)	1(6)	**1(1)**	***6(1,2,3,4,5,6)***	3(4,5,6)
	F	1(1)	1(1)	***6(1,2,3,4,5,6)***	1(1)	1(1)	1(1)	1(1)	**1(1)**	***6(1,2,3,4,5,6)***	4(1,2,4,5)
	M	1(1)	1(1)	***6(1,2,3,4,5,6)***	1(1)	1(1)	2(1,2)	1(1)	**1(1)**	***6(1,2,3,4,5,6)***	2(3,4)
	A	3(1,2,5)	3(1,2,5)	***6(1,2,3,4,5,6)***	3(1,2,5)	2(1,2)	3(1,2,3)	2(1,2)	**2(1,2)**	***6(1,2,3,4,5,6)***	1(4)
	M	4(1,2,3,5)	4(1,2,3,5)	***6(1,2,3,4,5,6)***	2(1,3)	1(1)	5(1,2,3,4,5)	6(1,2,3,4,5,6)	**3(1,2,3)**	***6(1,2,3,4,5,6)***	2(5,6)
	J	3(1,3,5)	4(1,2,3,5)	***5(1,2,3,5,6)***	1(1)	1(1)	5(1,2,3,4,5)	3(1,2,5)	**2(1,2)**	***5(1,2,3,5,6)***	2(1,2)
	J	3(1,2,6)	3(1,2,6)	***3(1,2,6)***	1(1)	1(1)	6(1,2,3,4,5,6)	2(1,2)	**2(1,2)**	***3(1,2,6)***	1(1)
	A	1(1)	1(1)	***1(1)***	2(1,3)	1(1)	6(1,2,3,4,5,6)	2(1,2)	**1(1)**	***1(1)***	2(1,2)
	S	1(1)	1(1)	***2(1,2)***	2(1,3)	1(1)	6(1,2,3,4,5,6)	2(1,2)	**4(1,2,3,4)**	***2(1,2)***	2(1,2)
	O	1(1)	1(1)	***3(1,2,3)***	4(1,3,4,6)	1(1)	6(1,2,3,4,5,6)	1(1)	**4(1,2,3,4)**	***3(1,2,3)***	1(2)
	N	2(1,2)	2(1,2)	***4(1,2,3,4)***	3(1,2,5)	2(1,2)	1(1)	1(1)	**1(1)**	***4(1,2,3,4)***	2(4,5)
	D	1(1)	1(1)	***5(1,2,3,4,5)***	3(1,5,6)	1(1)	2(1,2)	2(1,6)	**2(1,2)**	***5(1,2,3,4,5)***	3(1,2,3)
ELM	**J**	1(2)	1(2)	**3(1,2,6)**	2(1,6)	1(1)	1(6)	1(6)	1(1)	4(1,2,3,5)	3(4,5,6)
	F	1(1)	1(1)	**2(1,5)**	1(1)	1(1)	1(1)	1(1)	1(1)	2(1,4)	2(1,3)
	M	1(1)	1(1)	**2(2,3)**	1(1)	1(1)	2(1,2)	1(1)	1(1)	1(2)	5(1,2,4,5,6)
	A	1(5)	1(5)	**1(5)**	3(1,2,5)	2(1,2)	3(1,2,3)	2(1,2)	2(1,2)	1(3)	1(5)
	M	1(2)	1(2)	**1(2)**	2(1,3)	1(1)	5(1,2,3,4,5)	6(1,2,3,4,5,6)	3(1,2,3)	2(1,4)	2(1,5)
	J	1(1)	1(1)	**3(1,5,6)**	1(1)	1(1)	5(1,2,3,4,5)	3(1,2,5)	2(1,2)	2(1,5)	5(1,2,3,4,5)
	J	1(1)	1(1)	**2(1,2)**	1(1)	1(1)	6(1,2,3,4,5,6)	2(1,2)	2(1,2)	2(1,6)	4(1,3,4,6)
	A	1(1)	1(1)	**3(1,4,5)**	2(1,3)	1(1)	6(1,2,3,4,5,6)	2(1,2)	1(1)	2(1,6)	4(1,2,3,5)
	S	1(1)	1(1)	**3(1,5,6)**	2(1,3)	1(1)	6(1,2,3,4,5,6)	2(1,2)	4(1,2,3,4)	5(1,2,4,5,6)	3(2,5,6)
	O	1(1)	1(1)	**3(1,3,4)**	4(1,3,4,6)	1(1)	6(1,2,3,4,5,6)	1(1)	4(1,2,3,4)	5(1,2,3,4,5)	4(1,2,4,6)
	N	1(5)	1(5)	**3(1,4,5)**	3(1,2,5)	2(1,2)	1(1)	1(1)	1(1)	2(1,3)	3(1,2,3)
	D	1(2)	1(2)	**2(2,5)**	3(1,5,6)	1(1)	2(1,2)	2(1,6)	2(1,2)	3(1,2,5)	4(2,4,5,6)
AR		2(1,2)	2(1,2)	2(1,2)	**4(1,2,3,4)**	**4(1,2,3,4)**	***6(1,2,3,4,5,6)***	***6(1,2,3,4,5,6)***	***6(1,2,3,4,5,6)***	2(1,2)	2(1,2)
ELM		1(1)	**2(1,4)**	**6(1,2,3,4,5,6)**	**4(1,2,3,4)**	**4(1,2,3,4)**	**6(1,2,3,4,5,6)**	**6(1,2,3,4,5,6)**	**6(1,2,3,4,5,6)**	**6(1,2,3,4,5,6)**	**4(1,2,4,5,6)**

Table A3. Sobradinho.

Month		BIC	AIC	W-MSE	PACF	PACF -Sted.	MI	PMI	N-MRMCR -MI	GA	PSO
PAR	**J**	1(1)	1(1)	6(1,2,3,4,5,6)	1(1)	1(1)	**3(1,2,3)**	1(1)	***4(1,4,5,6)***	6(1,2,3,4,5,6)	2(3,6)
	F	3(1,3,5)	3(1,3,5)	6(1,2,3,4,5,6)	3(1,4,5)	1(1)	**4(1,2,4,5)**	1(1)	***6(1,2,3,4,5,6)***	6(1,2,3,4,5,6)	2(1,4)
	M	1(1)	1(1)	5(1,2,3,4,6)	1(1)	1(1)	**2(1,2)**	1(1)	***6(1,2,3,4,5,6)***	5(1,2,3,4,6)	3(1,2,5)
	A	1(1)	1(1)	6(1,2,3,4,5,6)	1(1)	1(1)	**5(1,2,3,4,6)**	1(5)	***6(1,2,3,4,5,6)***	6(1,2,3,4,5,6)	3(2,4,6)
	M	1(1)	2(1,3)	6(1,2,3,4,5,6)	1(1)	1(1)	**4(1,2,3,4)**	2(1,3)	***5(1,2,4,5,6)***	6(1,2,3,4,5,6)	3(1,3,4)
	J	5(1,2,3,5,6)	5(1,2,3,5,6)	6(1,2,3,4,5,6)	3(1,2,3)	3(1,2,3)	**6(1,2,3,4,5,6)**	6(1,2,3,4,5,6)	***6(1,2,3,4,5,6)***	6(1,2,3,4,5,6)	6(1,2,3,4,5,6)
	J	5(1,2,3,4,6)	6(1,2,3,4,5,6)	6(1,2,3,4,5,6)	1(1)	1(1)	**6(1,2,3,4,5,6)**	2(1,2)	***6(1,2,3,4,5,6)***	5(1,2,3,5,6)	2(1,3)
	A	1(1)	1(1)	5(1,2,3,4,5)	2(1,2)	2(1,2)	**6(1,2,3,4,5,6)**	3(1,5,4)	***6(1,2,3,4,5,6)***	5(1,2,3,4,5)	4(1,2,3,5)
	S	1(1)	1(1)	6(1,2,3,4,5,6)	2(1,2)	2(1,2)	**6(1,2,3,4,5,6)**	2(1,2)	***6(1,2,3,4,5,6)***	6(1,2,3,4,5,6)	5(1,2,3,4,5)
	O	1(1)	1(1)	4(1,2,3,4)	2(1,2)	2(1,2)	**6(1,2,3,4,5,6)**	3(1,2,6)	***6(1,2,3,4,5,6)***	4(1,2,3,4)	6(1,2,3,4,5,6)
	N	1(1)	2(1,3)	2(1,3)	1(1)	1(1)	**5(1,2,3,4,5)**	1(1)	***6(1,2,3,4,5,6)***	2(1,3)	1(1)
	D	1(1)	1(1)	5(1,2,3,4,5)	2(1,6)	1(1)	**2(1,2)**	1(1)	***6(1,2,3,4,5,6)***	5(1,2,3,4,5)	1(6)
ELM	**J**	1(2)	1(2)	**1(3)**	1(1)	1(1)	3(1,2,3)	1(1)	4(1,4,5,6)	1(3)	3(1,3,5)
	F	1(1)	1(1)	**3(1,2,6)**	3(1,4,5)	1(1)	4(1,2,4,5)	1(1)	6(1,2,3,4,5,6)	2(1,6)	4(1,2,5,6)
	M	1(1)	1(1)	**2(1,3)**	1(1)	1(1)	2(1,2)	1(1)	6(1,2,3,4,5,6)	3(1,5,6)	3(1,4,6)
	A	1(5)	1(5)	**3(1,2,4)**	1(1)	1(1)	5(1,2,3,4,6)	1(5)	6(1,2,3,4,5,6)	2(1,6)	5(1,2,3,5,6)
	M	1(2)	1(2)	**1(1)**	1(1)	1(1)	4(1,2,3,4)	2(1,3)	5(1,2,4,5,6)	2(1,3)	1(1)
	J	1(1)	1(1)	**2(1,4)**	3(1,2,3)	3(1,2,3)	6(1,2,3,4,5,6)	6(1,2,3,4,5,6)	6(1,2,3,4,5,6)	5(1,2,4,5,6)	3(1,3,6)
	J	1(1)	1(1)	**1(1)**	1(1)	1(1)	6(1,2,3,4,5,6)	2(1,2)	6(1,2,3,4,5,6)	1(1)	2(1,4)
	A	1(1)	1(1)	**1(1)**	2(1,2)	2(1,2)	6(1,2,3,4,5,6)	3(1,5,4)	6(1,2,3,4,5,6)	1(1)	2(1,2)
	S	1(1)	1(1)	**2(1,5)**	2(1,2)	2(1,2)	6(1,2,3,4,5,6)	2(1,2)	6(1,2,3,4,5,6)	2(1,2)	3(1,4,5)
	O	1(1)	1(1)	**5(1,2,4,5,6)**	2(1,2)	2(1,2)	6(1,2,3,4,5,6)	3(1,2,6)	6(1,2,3,4,5,6)	1(3)	4(1,4,5,6)
	N	1(5)	1(5)	**2(1,3)**	1(1)	1(1)	5(1,2,3,4,5)	1(1)	6(1,2,3,4,5,6)	1(3)	4(1,4,5,6)
	D	1(2)	1(2)	**4(2,3,5,6)**	2(1,6)	1(1)	2(1,2)	1(1)	6(1,2,3,4,5,6)	1(2)	5(1,2,3,4,5)
AR		2(1,3)	2(1,3)	2(1,3)	2(1,3)	3(1,3,4)	**6(1,2,3,4,5,6)**	**6(1,2,3,4,5,6)**	**6(1,2,3,4,5,6)**	2(1,3)	2(1,3)
ELM		1(1)	**2(1,4)**	6(1,2,3,4,5,6)	1(1)	3(1,3,4)	6(1,2,3,4,5,6)	6(1,2,3,4,5,6)	6(1,2,3,4,5,6)	6(1,2,3,4,5,6)	2(1,4)

Table A4. Agua Vermelha.

Month		BIC	AIC	W-MSE	PACF	PACF -Sted.	MI	PMI	N-MRMCR -MI	GA	PSO
PAR	**J**	1(1)	1(1)	6(1,2,3,4,5,6)	2(1,5)	**1(1)**	3(1,5,6)	1(1)	1(1)	***5(1,3,4,5,6)***	3(3,4,6)
	F	1(1)	1(1)	6(1,2,3,4,5,6)	2(1,6)	**1(1)**	1(1)	1(1)	1(1)	***6(1,2,3,4,5,6)***	2(5,6)
	M	1(1)	1(1)	6(1,2,3,4,5,6)	2(1,6)	**1(1)**	2(1,2)	1(1)	1(1)	***6(1,2,3,4,5,6)***	3(3,5,6)
	A	3(1,2,6)	3(1,2,6)	4(1,2,5,6)	2(1,2)	**2(1,2)**	3(1,2,4)	1(1)	2(1,2)	***4(1,2,5,6)***	2(2,3)
	M	4(1,2,3,5)	4(1,2,3,5)	5(1,2,3,4,5)	3(1,2,3)	**3(1,2,3)**	5(1,2,3,4,5)	2(1,2)	3(1,2,3)	***5(1,2,3,4,5)***	3(2,5,6)
	J	1(1)	1(1)	5(1,2,4,5,6)	1(1)	**1(1)**	5(1,2,3,4,5)	6(1,2,3,4,5,6)	2(1,2)	***5(1,2,4,5,6)***	4(2,3,4,5)
	J	3(1,2,3)	3(1,2,3)	6(1,2,3,4,5,6)	2(1,2)	**2(1,2)**	6(1,2,3,4,5,6)	2(1,2)	2(1,2)	***6(1,2,3,4,5,6)***	1(3)
	A	1(1)	5(1,2,3,4,6)	6(1,2,3,4,5,6)	1(1)	**1(1)**	6(1,2,3,4,5,6)	6(1,2,3,4,5,6)	1(1)	***6(1,2,3,4,5,6)***	5(2,3,4,5,6)
	S	2(1,3)	2(1,3)	6(1,2,3,4,5,6)	2(1,2)	**2(1,2)**	6(1,2,3,4,5,6)	1(1)	4(1,2,3,4)	***6(1,2,3,4,5,6)***	1(1)
	O	2(1,3)	3(1,3,6)	6(1,2,3,4,5,6)	3(1,2,3)	**3(1,2,3)**	6(1,2,3,4,5,6)	6(1,2,3,4,5,6)	4(1,2,3,4)	***6(1,2,3,4,5,6)***	1(1)
	N	1(1)	2(1,2)	6(1,2,3,4,5,6)	1(1)	**1(1)**	6(1,2,3,4,5,6)	1(1)	1(1)	***6(1,2,3,4,5,6)***	4(1,2,5,6)
	D	1(1)	1(1)	6(1,2,3,4,5,6)	1(1)	**1(1)**	5(1,2,3,4,5)	1(1)	2(1,2)	***6(1,2,3,4,5,6)***	1(4)
ELM	**J**	1(2)	1(2)	**2(1,6)**	2(1,5)	1(1)	3(1,5,6)	1(1)	1(1)	5(1,2,3,4,6)	4(1,2,4,6)
	F	1(1)	1(1)	**1(1)**	2(1,6)	1(1)	1(1)	1(1)	1(1)	4(1,2,3,6)	4(1,2,4,5)
	M	1(1)	1(1)	**2(1,2)**	2(1,6)	1(1)	2(1,2)	1(1)	1(1)	2(1,2)	6(1,2,3,4,5,6)
	A	1(5)	1(5)	**2(1,2)**	2(1,2)	2(1,2)	3(1,2,4)	1(1)	2(1,2)	2(1,2)	4(1,2,3,5)
	M	1(2)	1(2)	**1(1)**	3(1,2,3)	3(1,2,3)	5(1,2,3,4,5)	2(1,2)	3(1,2,3)	4(1,2,4,6)	4(2,3,5,6)
	J	1(1)	1(1)	**3(1,2,6)**	1(1)	1(1)	5(1,2,3,4,5)	6(1,2,3,4,5,6)	2(1,2)	2(2,5)	3(1,2,4)
	J	1(1)	1(1)	**2(1,2)**	2(1,2)	2(1,2)	6(1,2,3,4,5,6)	2(1,2)	2(1,2)	2(1,2)	3(1,2,5)
	A	1(1)	1(1)	**3(1,2,5)**	1(1)	1(1)	6(1,2,3,4,5,6)	6(1,2,3,4,5,6)	1(1)	5(1,2,3,4,5)	2(1,2)
	S	1(1)	1(1)	**2(4,6)**	2(1,2)	2(1,2)	6(1,2,3,4,5,6)	1(1)	4(1,2,3,4)	2(2,6)	2(5,6)
	O	1(1)	1(1)	**2(1,6)**	3(1,2,3)	3(1,2,3)	6(1,2,3,4,5,6)	6(1,2,3,4,5,6)	4(1,2,3,4)	3(1,3,4)	3(1,2,5)
	N	1(5)	1(5)	**1(6)**	1(1)	1(1)	6(1,2,3,4,5,6)	1(1)	1(1)	2(2,5)	4(2,3,5,6)
	D	1(2)	1(2)	**3(2,3,4)**	1(1)	1(1)	5(1,2,3,4,5)	1(1)	2(1,2)	1(2)	2(2,6)
AR		2(1,2)	2(1,2)	2(1,2)	**3(1,2,3)**	**3(1,2,3)**	6(1,2,3,4,5,6)	***2(1,3)***	6(1,2,3,4,5,6)	6(1,2,3,4,5,6)	6(1,2,3,4,5,6)
ELM		1(1)	**2(1,4)**	6(1,2,3,4,5,6)	3(1,2,3)	3(1,2,3)	6(1,2,3,4,5,6)	2(1,3)	6(1,2,3,4,5,6)	6(1,2,3,4,5,6)	4(1,2,3,4)

Table A5. Passo Real.

Month		BIC	AIC	W-MSE	PACF	PACF -Sted.	MI	PMI	N-MRMCR -MI	GA	PSO
PAR	**J**	2(1,2)	2(1,2)	6(1,2,3,4,5,6)	3(1,5,6)	1(1)	**3(1,2,4)**	1(1)	6(1,2,3,4,5,6)	***6(1,2,3,4,5,6)***	5(1,2,4,5,6)
	F	2(1,3)	3(1,3,5)	5(1,3,4,5,6)	1(1)	1(1)	**3(1,2,3)**	1(1)	6(1,2,3,4,5,6)	***5(1,3,4,5,6)***	3(2,5,6)
	M	1(1)	2(1,6)	5(1,2,3,4,6)	2(1,2)	2(1,2)	**6(1,2,3,4,5,6)**	1(1)	6(1,2,3,4,5,6)	***5(1,2,3,4,6)***	2(2,5)
	A	1(2)	1(2)	6(1,2,3,4,5,6)	3(1,2,4)	2(1,2)	**5(1,2,3,4,5)**	3(1,2,3)	6(1,2,3,4,5,6)	***6(1,2,3,4,5,6)***	3(1,5,6)
	M	1(1)	2(1,5)	4(1,3,5,6)	5(1,2,3,4,6)	4(1,2,3,4)	**6(1,2,3,4,5,6)**	2(1,6)	6(1,2,3,4,5,6)	***4(1,3,5,6)***	2(2,3)
	J	1(1)	2(1,2)	6(1,2,3,4,5,6)	6(1,2,3,4,5,6)	6(1,2,3,4,5,6)	**4(1,2,3,5)**	3(1,2,5)	6(1,2,3,4,5,6)	***6(1,2,3,4,5,6)***	3(1,3,5)
	J	3(1,2,4)	3(1,2,4)	6(1,2,3,4,5,6)	6(1,2,3,4,5,6)	6(1,2,3,4,5,6)	**5(1,2,3,4,6)**	2(1,3)	6(1,2,3,4,5,6)	***6(1,2,3,4,5,6)***	4(3,4,5,6)
	A	1(1)	3(1,2,5)	6(1,2,3,4,5,6)	6(1,2,3,4,5,6)	6(1,2,3,4,5,6)	**5(1,2,3,4,5)**	2(1,5)	6(1,2,3,4,5,6)	***6(1,2,3,4,5,6)***	3(2,3,6)
	S	1(1)	1(1)	6(1,2,3,4,5,6)	6(1,2,3,4,5,6)	6(1,2,3,4,5,6)	**2(1,3)**	2(1,3)	6(1,2,3,4,5,6)	***6(1,2,3,4,5,6)***	3(2,4,5)
	O	1(1)	3(1,2,4)	5(1,2,4,5,6)	6(1,2,3,4,5,6)	6(1,2,3,4,5,6)	**2(1,4)**	2(1,4)	6(1,2,3,4,5,6)	***5(1,2,4,5,6)***	2(3,4)
	N	3(1,2,3)	4(1,2,3,5)	6(1,2,3,4,5,6)	6(1,2,3,4,5,6)	6(1,2,3,4,5,6)	**2(1,3)**	2(1,3)	6(1,2,3,4,5,6)	***6(1,2,3,4,5,6)***	2(5,6)
	D	1(1)	2(1,4)	6(1,2,3,4,5,6)	5(1,2,3,4,6)	4(1,2,3,4)	**3(1,2,4)**	1(1)	6(1,2,3,4,5,6)	***5(1,3,4,5,6)***	1(1)
ELM	**J**	1(2)	1(2)	**1(2)**	3(1,5,6)	1(1)	3(1,2,4)	1(1)	6(1,2,3,4,5,6)	2(1,4)	3(1,3,4)
	F	1(1)	1(1)	**1(1)**	1(1)	1(1)	3(1,2,3)	1(1)	6(1,2,3,4,5,6)	2(1,2)	5(1,2,3,4,6)
	M	1(1)	1(1)	**2(3,6)**	2(1,2)	2(1,2)	6(1,2,3,4,5,6)	1(1)	6(1,2,3,4,5,6)	2(1,2)	3(1,3,6)
	A	1(5)	1(5)	**1(5)**	3(1,2,4)	2(1,2)	5(1,2,3,4,5)	3(1,2,3)	6(1,2,3,4,5,6)	2(1,2)	3(1,3,4)
	M	1(2)	1(2)	**1(2)**	5(1,2,3,4,6)	4(1,2,3,4)	6(1,2,3,4,5,6)	2(1,6)	6(1,2,3,4,5,6)	1(4)	4(1,2,3,6)
	J	1(1)	1(1)	**1(1)**	6(1,2,3,4,5,6)	6(1,2,3,4,5,6)	4(1,2,3,5)	3(1,2,5)	6(1,2,3,4,5,6)	2(2,5)	1(1)
	J	1(1)	1(1)	**1(1)**	6(1,2,3,4,5,6)	6(1,2,3,4,5,6)	5(1,2,3,4,6)	2(1,3)	6(1,2,3,4,5,6)	3(2,3,6)	2(1,5)
	A	1(1)	1(1)	**1(1)**	6(1,2,3,4,5,6)	6(1,2,3,4,5,6)	5(1,2,3,4,5)	2(1,5)	6(1,2,3,4,5,6)	3(3,4,5)	3(2,3,5)
	S	1(1)	1(1)	**1(1)**	6(1,2,3,4,5,6)	6(1,2,3,4,5,6)	2(1,3)	2(1,3)	6(1,2,3,4,5,6)	4(1,2,4,5)	3(2,3,5)
	O	1(1)	1(1)	**1(1)**	6(1,2,3,4,5,6)	6(1,2,3,4,5,6)	2(1,4)	2(1,4)	6(1,2,3,4,5,6)	3(1,4,5)	3(2,4,6)
	N	1(5)	1(5)	**2(2,5)**	6(1,2,3,4,5,6)	6(1,2,3,4,5,6)	2(1,3)	2(1,3)	6(1,2,3,4,5,6)	3(1,2,3)	1(1)
	D	1(2)	1(2)	**1(2)**	5(1,2,3,4,6)	4(1,2,3,4)	3(1,2,4)	1(1)	6(1,2,3,4,5,6)	3(1,4,5)	2(1,3)
AR		2(1,2)	2(1,2)	2(1,2)	**3(1,2,3)**	**3(1,2,3)**	***6(1,2,3,4,5,6)***	***6(1,2,3,4,5,6)***	***6(1,2,3,4,5,6)***	2(1,2)	2(1,2)
ELM		1(1)	2(1,4)	6(1,2,3,4,5,6)	**3(1,2,3)**	**3(1,2,3)**	6(1,2,3,4,5,6)	6(1,2,3,4,5,6)	6(1,2,3,4,5,6)	6(1,2,3,4,5,6)	4(1,2,3,6)

References

1. IHA—International Hydropower Association. Hydropower Status Report: Sector Trends and Insights. 2020. Available online: https://www.hydropower.org/publications/2020-hydropower-status-report (accessed on 15 May 2020).
2. Siqueira, H.V.; Boccato, L.; Luna, I.; Attux, R.; Lyra, C. Performance analysis of unorganized machines in streamflow forecasting of Brazilian plants. *Appl. Soft Comput.* **2018**, *68*, 494–506. [CrossRef]
3. Zhu, S.; Zhou, J.; Ye, L.; Meng, C. Streamflow estimation by support vector machine coupled with different methods of time series decomposition in the upper reaches of Yangtze River, China. *Environ. Earth Sci.* **2016**, *75*, 531. [CrossRef]
4. Dilini, W.; Attygalle, D.; Hansen, L.L.; Nandalal, K.W. Ensemble Forecast for monthly Reservoir Inflow; A Dynamic Neural Network Approach. In Proceedings of the 4th Annual International Conference on Operations Research and Statistics (ORS 2016), Global Science and Technology Forum, Singapore, 18–19 January 2016; pp. 84–90.
5. Fouad, G.; Loáiciga, H.A. Independent variable selection for regression modeling of the flow duration curve for ungauged basins in the United States. *J. Hydrol.* **2020**, *587*, 124975. [CrossRef]
6. Arsenault, R.; Côté, P. Analysis of the effects of biases in ensemble streamflow prediction (ESP) forecasts on electricity production in hydropower reservoir management. *Hydrol. Earth Syst. Sci.* **2019**, *23*, 2735–2750. [CrossRef]
7. Stojković, M.; Kostić, S.; Prohaska, S.; Plavsic, J.; Tripković, V. A new approach for trend assessment of annual streamflows: A case study of hydropower plants in Serbia. *Water Resour. Manag.* **2017**, *31*, 1089–1103. [CrossRef]
8. Hailegeorgis, T.T.; Alfredsen, K. Regional statistical and precipitation-runoff modelling for ecological applications: prediction of hourly streamflow in regulated rivers and ungauged basins. *River Res. Appl.* **2016**, *33*, 233–248. [CrossRef]
9. Hernandez-Ambato, J.; Asqui-Santillan, G.; Arellano, A.; Cunalata, C. Multistep-ahead Streamflow and Reservoir Level Prediction Using ANNs for Production Planning in Hydroelectric Stations. In Proceedings of the 16th IEEE International Conference on Machine Learning and Applications (ICMLA 2017), Institute of Electrical and Electronics Engineers (IEEE), Cancun, Mexico, 18–21 December 2017; pp. 479–484.
10. Yaseen, Z.M.; Jaafar, O.; Deo, R.; Kisi, O.; Adamowski, J.; Quilty, J.; El-Shafie, A. Stream-flow forecasting using extreme learning machines: A case study in a semi-arid region in Iraq. *J. Hydrol.* **2016**, *542*, 603–614. [CrossRef]
11. Maceira, M.E.P.; Damázio, J.M. Use of the PAR (p) model in the stochastic dual dynamic programming optimization scheme used in the operation planning of the brazilian hydropower system. *Probab. Eng. Inf. Sci.* **2005**, *20*, 143–156. [CrossRef]
12. Siqueira, H.V.; Boccato, L.; Attux, R.; Lyra, C. Unorganized machines for seasonal streamflow series forecasting. *Int. J. Neural Syst.* **2014**, *24*, 1430009. [CrossRef] [PubMed]
13. Munera, S.; Amigo, J.M.; Aleixos, N.; Talens, P.; Cubero, S.; Blasco, J. Potential of VIS-NIR hyperspectral imaging and chemometric methods to identify similar cultivars of nectarine. *Food Control.* **2018**, *86*, 1–10. [CrossRef]
14. Yan, L.; Liu, C.; Qu, H.; Liu, W.; Zhang, Y.; Yang, J.; Zheng, L. Discrimination and measurements of three flavonols with similar structure using terahertz spectroscopy and chemometrics. *J. Infrared Millim. Terahertz Waves* **2018**, *39*, 492–504. [CrossRef]
15. Moon, Y.I.; Rajagopalan, B.; Lallid, U. Estimation of mutual information using kernel density estimators. *Phys. Rev. E* **1995**, *52*, 2318–2321. [CrossRef] [PubMed]
16. Crone, S.F.; Kourentzes, N. Feature selection for time series prediction—A combined filter and wrapper approach for neural networks. *Neurocomputing* **2010**, *73*, 1923–1936. [CrossRef]
17. Guyon, I.; Elisseeff, A. An introduction to variable and feature selection. *J. Mach. Learn. Res.* **2003**, *3*, 1157–1182. [CrossRef]
18. Wang, A.; Xie, L. Technology using near infrared spectroscopic and multivariate analysis to determine the soluble solids content of citrus fruit. *J. Food Eng.* **2014**, *143*, 17–24. [CrossRef]

19. Yin, L.; Zhou, J.; Chen, D.; Han, T.; Zheng, B.; Younis, A.; Shao, Q. A review of the application of near-infrared spectroscopy to rare traditional Chinese medicine. *Spectrochim. Acta Part. A Mol. Biomol. Spectrosc.* **2019**, *221*, 117208. [CrossRef]
20. Harrell, F.E. *Regression Modeling Strategies*; Springer Science and Business Media LLC: New York, NY, USA, 2001.
21. Tsakiris, G.; Nalbantis, I.; Cavadias, G. Regionalization of low flows based on canonical correlation analysis. *Adv. Water Resour.* **2011**, *34*, 865–872. [CrossRef]
22. Li, X.; Liu, Z.; Lin, H.; Wang, G.; Sun, H.; Long, J.; Zhang, M. Estimating the growing stem volume of chinese pine and larch plantations based on fused optical data using an improved variable screening method and stacking algorithm. *Remote. Sens.* **2020**, *12*, 871. [CrossRef]
23. Bonah, E.; Huang, X.; Aheto, J.H.; Yi, R.; Yu, S.; Tu, H. Comparison of variable selection algorithms on vis-NIR hyperspectral imaging spectra for quantitative monitoring and visualization of bacterial foodborne pathogens in fresh pork muscles. *Infrared Phys. Technol.* **2020**, *107*, 103327. [CrossRef]
24. Xiong, Y.; Zhang, R.; Zhang, F.; Yang, W.; Kang, Q.; Chen, W.; Du, Y. A spectra partition algorithm based on spectral clustering for interval variable selection. *Infrared Phys. Technol.* **2020**, *105*, 103259. [CrossRef]
25. Speiser, J.L.; Miller, M.A.; Tooze, J.; Ip, E.H. A comparison of random forest variable selection methods for classification prediction modeling. *Expert Syst. Appl.* **2019**, *134*, 93–101. [CrossRef]
26. Rendall, R.; Pereira, A.C.; Reis, M. An extended comparison study of large scale datadriven prediction methods based on variable selection, latent variables, penalized regression and machine learning. *Comput. Aided Chem. Eng.* **2016**, *38*, 1629–1634. [CrossRef]
27. Marcjasz, G.; Uniejewski, B.; Weron, R. Beating the naïve—Combining LASSO with naïve intraday electricity price forecasts. *Energies* **2020**, *13*, 1667. [CrossRef]
28. Santi, V.M.A.; Notodiputro, K.; Sartono, B. Variable selection methods applied to the mathematics scores of Indonesian students based on convex penalized likelihood. *J. Phys. Conf. Ser.* **2019**, *1402*, 077096. [CrossRef]
29. Karim, N.; Reid, C.M.; Tran, L.; Cochrane, A.; Billah, B. Variable selection methods for multiple regressions influence the parsimony of risk prediction models for cardiac surgery. *J. Thorac. Cardiovasc. Surg.* **2017**, *153*, 1128–1135.e3. [CrossRef]
30. Kim, D.; Kang, S. Effect of irrelevant variables on faulty wafer detection in semiconductor manufacturing. *Energies* **2019**, *12*, 2530. [CrossRef]
31. Furmańczyk, K.; Rejchel, W. Prediction and variable selection in high-dimensional misspecified binary classification. *Entropy* **2020**, *22*, 543. [CrossRef]
32. Tutkun, N.A.; Atilgan, Y.K. Visual research on the trustability of classical variable selection methods in Cox regression. *Hacet. J. Math. Stat.* **2020**, *49*, 1–18. [CrossRef]
33. Mehmood, T.; Saebø, S.; Liland, K.H. Comparison of variable selection methods in partial least squares regression. *J. Chemom.* **2020**, *34*, e3226. [CrossRef]
34. McGee, M.; Yaffee, R.A. Comparison of Variable Selection Methods for Forecasting from Short Time Series. In Proceedings of the 6th IEEE International Conference on Data Science and Advanced Analytics (DSAA 2019), Institute of Electrical and Electronics Engineers (IEEE), Washington, DC, USA, 5–8 October 2019; pp. 531–540.
35. Seo, H.S. Unified methods for variable selection and outlier detection in a linear regression. *Commun. Stat. Appl. Methods* **2019**, *26*, 575–582. [CrossRef]
36. Dong, W.; Yang, Q.; Fang, X. Multi-Step ahead wind power generation prediction based on hybrid machine learning techniques. *Energies* **2018**, *11*, 1975. [CrossRef]
37. Sigauke, C.; Nemukula, M.M.; Maposa, D. Probabilistic hourly load forecasting using additive quantile regression models. *Energies* **2018**, *11*, 2208. [CrossRef]
38. Wang, H.; Sun, J.; Sun, J.; Wang, J. Using random forests to select optimal input variables for short-term wind speed forecasting models. *Energies* **2017**, *10*, 1522. [CrossRef]
39. Taormina, R.; Ahmadi, M.H. Data-driven input variable selection for rainfall–runoff modeling using binary-coded particle swarm optimization and Extreme Learning Machines. *J. Hydrol.* **2015**, *529*, 1617–1632. [CrossRef]
40. Taormina, R.; Ahmadi, M.H.; Sivakumar, B. Neural network river forecasting through baseflow separation and binary-coded swarm optimization. *J. Hydrol.* **2015**, *529*, 1788–1797. [CrossRef]
41. Cui, X.; Jiang, M. Chaotic time series prediction based on binary particle swarm optimization. *AASRI Proc.* **2012**, *1*, 377–383. [CrossRef]

42. Silva, N.; Siqueira, I.; Okida, S.; Stevan, S.L.; Siqueira, H.V. Neural networks for predicting prices of sugarcane derivatives. *Sugar Tech.* **2018**, *21*, 514–523. [CrossRef]
43. Siqueira, H.V.; Boccato, L.; Attux, R.; Lyra, C. Echo state networks and extreme learning machines: A comparative study on seasonal streamflow series prediction. In *Computer Vision*; Springer Science and Business Media LLC: Heidelberg/Berlin, Germany, 2012; Volume 7664, pp. 491–500.
44. Siqueira, H.V.; Boccato, L.; Attux, R.; Filho, C.L. Echo state networks for seasonal streamflow series forecasting. In *Computer Vision*; Springer Science and Business Media LLC: Berlin/Heidelberg, Germany, 2012; Volume 7435, pp. 226–236.
45. Siqueira, H.V.; Boccato, L.; Attux, R.; Filho, C.L. Echo State networks in seasonal streamflow series prediction. *Learn. Nonlinear Model* **2012**, *10*, 181–191. [CrossRef]
46. Kachba, Y.R.; Chiroli, D.M.D.G.; Belotti, J.T.; Alves, T.A.; Tadano, Y.D.S.; Siqueira, H.V. Artificial neural networks to estimate the influence of vehicular emission variables on morbidity and mortality in the largest metropolis in South America. *Sustainability* **2020**, *12*, 2621. [CrossRef]
47. Puma-Villanueva, W.; Dos Santos, E.; Von Zuben, F. Data partition and variable selection for time series prediction using wrappers. In Proceedings of the 2006 IEEE International Joint Conference on Neural Network Proceedings, Institute of Electrical and Electronics Engineers (IEEE), Vancouver, BC, Canada, 16–21 July 2006; pp. 4740–4747.
48. Yu, L.; Liu, H. Efficient feature selection via analysis of relevance and redundancy. *J. Mach. Learn. Res.* **2004**, *5*, 1205–1224.
49. Hyvärinen, A.; Karhunen, J.; Oja, E. *Independent Component Analysis*; John Wiley and Sons: New York, NY, USA, 2001; ISBN 978-0-471-40540-5.
50. Geurts, M.; Box, G.E.P.; Jenkins, G.M. *Time Series Analysis: Forecasting and Control*, 5th ed.; John Wiley and Sons: Hoboken, NJ, USA, 2016; ISBN 978-1-118-67502-1.
51. Quenouille, M.H. Approximate tests of correlation in time-series. *J. R. Stat. Soc. Ser. B* **1949**, *11*, 68–84. [CrossRef]
52. Stedinger, J.R. *Report on the Evaluation of CEPEL's PAR Models, Technical Report*; School of Civil and Environmental Engineering—Cornell University, Ithaca: New York, NY, USA, 2001.
53. Bonnlander, V.; Weigend, A.S. Selecting Input Variables Using Mutual Information and Nonparametric Density Estimation. In Proceedings of the 1994 International Symposium on Artificial Neural Networks (ISANN'94), National Cheng Kung University, Taiwan, China, 28 June–2 July 1994; pp. 42–50.
54. Luna, I.; Soares, S.; Ballini, R. Partial Mutual Information Criterion for Modelling Time Series Via Neural Networks. In Proceedings of the 11th Information Processing and Management of Uncertainty in Knowledge-Based System (IPMU 2006), Université Pierre et Marie Curie, Paris, France, 2–7 July 2006; pp. 2012–2019.
55. Bowden, G.J.; Dandy, G.; Maier, H.R. Input determination for neural network models in water resources applications. Part 1—Background and methodology. *J. Hydrol.* **2005**, *301*, 75–92. [CrossRef]
56. Akaho, S. Conditionally independent component analysis for supervised feature extraction. *Neurocomputing* **2002**, *49*, 139–150. [CrossRef]
57. Luna, I.; Ballini, R. Top-down strategies based on adaptive fuzzy rule-based systems for daily time series forecasting. *Int. J. Forecast.* **2011**, *27*, 708–724. [CrossRef]
58. Sharma, A. Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: Part 1—A strategy for system predictor identification. *J. Hydrol.* **2000**, *239*, 232–239. [CrossRef]
59. Specht, D.F. A general regression neural network. *IEEE Trans. Neural Netw.* **1991**, *2*, 568–576. [CrossRef]
60. Bowden, G.J. Forecasting Water Resources Variables Using Artificial Neural Networks. Ph.D. Thesis, University of Adelaide, Adelaide, Australia, February 2003.
61. Scott, D.W. *Multivariate Density Estimation: Theory, Practice, And Visualization*; John Wiley and Sons: New York, NY, USA, 1992; ISBN 978-0-471-54770-9.
62. Che, J.; Yang, Y.; Li, L.; Bai, X.; Zhang, S.; Deng, C.; Fowler, J.E. Maximum relevance minimum common redundancy feature selection for nonlinear data. *Inf. Sci.* **2017**, 68–86. [CrossRef]
63. Kohavi, R.; John, G.H. Wrappers for feature subset selection. *Artif. Intell.* **1997**, *97*, 273–324. [CrossRef]
64. McLeod, A.I. Diagnostic checking of periodic autoregression models with application. *J. Time Ser. Anal.* **1994**, *15*, 221–233. [CrossRef]
65. Schwarz, G. Estimating the Dimension of a Model. *Ann. Stat.* **1978**, *6*, 461–464. [CrossRef]

66. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control.* **1974**, *19*, 716–723. [CrossRef]
67. Kennedy, J.; Eberhart, R. Particle swarm optimization. In Proceedings of the ICNN'95—International Conference on Neural Networks, Institute of Electrical and Electronics Engineers (IEEE), Perth, Australia, 27 November–1 December 1995; Volume 4, pp. 1942–1948.
68. Santana, C.J.; Macedo, M.; Siqueira, H.V.; Gokhale, A.A.; Bastos-Filho, C.J.A. A novel binary artificial bee colony algorithm. *Futur. Gener. Comput. Syst.* **2019**, *98*, 180–196. [CrossRef]
69. Siqueira, H.; Santana, C.; Macedo, M.; Figueiredo, E.; Gokhale, A.; Bastos-Filho, C. Simplified binary cat swarm optimization. *Integr. Comput. Eng.* **2020**, 1–15. [CrossRef]
70. Holland, J.H. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*, 1st ed.; MIT Press: Cambridge, MA, USA, 1992; ISBN 978-0-262-08213-6.
71. Kennedy, J.; Eberhart, R.C. A discrete binary version of the particle swarm algorithm. In Proceedings of the 1997 IEEE International Conference on Systems, Man, and Cybernetics, Computational Cybernetics and Simulation, Institute of Electrical and Electronics Engineers (IEEE), Orlando, FL, USA, 12–15 October 1997; Volume 5, pp. 4104–4108.
72. EPE—Energy Research Company (in Portuguese: Empresa de Pesquisa Energética). 2019; Brazilian National Energy Balance 2019 (Base Year 2012). Available online: https://www.epe.gov.br/en/publications/publications/brazilian-energy-balance (accessed on 15 May 2020).
73. Sacchi, R.; Ozturk, M.C.; Principe, J.C.; Carneiro, A.A.F.M.; Da Silva, I.N. Water Inflow Forecasting using the Echo State Network: A Brazilian Case Study. In Proceedings of the 2007 International Joint Conference on Neural Networks, Institute of Electrical and Electronics Engineers (IEEE), Orlando, FL, USA, 12–17 August 2007; pp. 2403–2408.
74. ONS—Electric System Operator—Brazil (in Portuguese: Operador Nacional do Sistema Elétrico). 2020. Dados Hidrológicos/Vazões. Available online: http://www.ons.org.br/Paginas/resultados-da-operacao/historico-da-operacao/dados_hidrologicos_vazoes.aspx (accessed on 1 May 2020).
75. Vecchia, A.V. Maximum likelihood estimation for periodic autoregressive moving average models. *Technometrics* **1985**, *27*, 375–384. [CrossRef]
76. Hipel, K.W.; McLeod, A.I. *Time Series Modelling of Water Resources and Environmental Systems*, 1st ed.; Elsevier: New York, NY, USA, 1994; ISBN 978-0-444-89270-6.
77. Huang, G.B.; Zhu, Q.Y.; Siew, C.K. Extreme learning machine: Theory and applications. *Neurocomputing* **2006**, *70*, 489–501. [CrossRef]
78. Bartlett, P.L. The sample complexity of pattern classification with neural networks: The size of the weights is more important than the size of the network. *IEEE Trans. Inf. Theory* **1998**, *44*, 525–536. [CrossRef]
79. Huang, G.; Huang, G.B.; Song, S.; You, K. Trends in extreme learning machines: A review. *Neural Netw.* **2015**, *61*, 32–48. [CrossRef]
80. Siqueira, H.; Luna, I. Performance comparison of feedforward neural networks applied to stream flow series forecasting. *Math. Eng. Sci. Aerosp.* **2019**, *10*, 41–53.
81. Kowalczyk-Juśko, A.; Pochwatka, P.; Zaborowicz, M.; Czekała, W.; Mazurkiewicz, J.; Mazur, A.; Janczak, D.; Marczuk, A.; Dach, J. Energy value estimation of silages for substrate in biogas plants using an artificial neural network. *Energy* **2020**, *202*, 117729. [CrossRef]

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).