ARTICLE IN PRESS

Available at www.sciencedirect.com

INFORMATION PROCESSING IN AGRICULTURE xxx (xxxx) xxx

journal homepage: www.keaipublishing.com/en/journals/information-processing-in-agriculture/

A methodology for coffee price forecasting based on extreme learning machines

Carolina Deina [a], *Matheus Henrique do Amaral Prates* [b], *Carlos Henrique Rodrigues Alves* [c], *Marcella Scoczynski Ribeiro Martins* [b], *Flavio Trojan* [b,c], *Sergio Luiz Stevan Jr.* [b], *Hugo Valadares Siqueira* [b,c,*]

[a] *Department of Production Engineering, Federal University of Rio Grande do Sul, 90.035-190, Brazil*
[b] *Department of Electronics, Federal University of Technology, Parana 84017-220, Brazil*
[c] *Graduate Program in Production Engineering, Federal University of Technology, Parana 84017-220, Brazil*

A R T I C L E I N F O

Article history:
Received 15 May 2020
Received in revised form
6 July 2021
Accepted 22 July 2021
Available online xxxx

Keyword:
Coffee price forecasting
Linear models
Artificial neural networks
Computational intelligence

A B S T R A C T

This work introduces a methodology to estimate coffee prices based on the use of Extreme Learning Machines. The process is initiated by identifying the presence of nonstationary components, like seasonality and trend. These components are withdrawn if they are found. Next, the temporal lags are selected based on the response of the Partial Autocorrelation Function filter. As predictors, we address the following models: Exponential Smoothing (ES), Autoregressive (AR) and Autoregressive Integrated and Moving Average (ARIMA) models, Multilayer Perceptron (MLP) and Extreme Learning Machines (ELMs) neural networks. The computational results based on three error metrics and two coffee types (Arabica and Robusta) showed that the neural networks, especially the ELM, can reach higher performance levels than the other models. The methodology, which presents preprocessing stages, lag selection, and use of ELM, is a novelty that contributes to the coffee prices forecasting field.

1. Introduction

Coffee is a delicacy not only appreciated in countries that have a tradition in its trade. Due to the importance of its commercialization, the coffee industry is expressive of the world economy [1]. According to the newest information provided by the International Coffee Organization, in 2021, the crop consumption for coffee from Abril 2019 to March 2020 was 164.2 million 60-kg bags in the world and the preview, and was projected an increase of 1.3% for the next year [1].

Domestic coffee consumption and exportation have moved the international trade balance. However, coffee exports are subject to price fluctuations, which brings risks to farmers, importers, marketing institutions, and consumers [2]. According to [3,4], this is due to volatility caused by changes in world market price trends.

The same authors state that the methodologies used to predict commodity prices are increasingly important, despite the intrinsic difficulties related to this topic. When the results generated are consistent, they aid not only the decision-making process, but also the whole production chain [5]. As

* Corresponding author at: Department of Electronics into Federal University of Technology – Parana (UTFPR), Rua Doutor Washington Subtil Chueire, 330, Ponta Grossa 84017-220, Brazil.
E-mail address: hugosiqueira@utfpr.edu.br (H.V. Siqueira).
Peer review under responsibility of China Agricultural University.
https://doi.org/10.1016/j.inpa.2021.07.003

a result, it becomes relevant to investigate new strategies and tools to overcome this problem.

Due to the importance of the subject, relevant studies have been developed on coffee price forecasting in different places in the world. In [6], the authors estimated linear and nonlinear error correction models for the spot prices of four different coffee types: Unwashed Arabica (Brazil), Colombian Mild Arabica (Colombia), other Mild Arabica (from other Latin American countries), and Robusta (from Africa and Asia). In [7–9], the authors used linear models like exponential approaches (Holt's model, Brown's linear trend model, Winter's seasonal exponential smoothing method, Holt-Winters, ARIMA, GARCH), Artificial Neural Networks, and Markov Chains for price and export forecasting of Indian coffee.

In [10], an Artificial Neural Network (ANN) and a Support Vector Machine (SVM) were used for the price prediction of a leading green coffee export company in the Indian stock market. The same authors developed a cognitive decision support system for the Indian green coffee supply chain based on ARIMA models and Least Square SVR to solve the same problem [11]. Besides that, versions of the Autoregressive Integrated Moving Average (ARIMA) were addressed in [11–13], among other linear approaches for coffee price forecasting.

In [14], the proposers used quantile regression models to identify the Granger causalities of some trading activities that influence the prices of commodities in the United States. In [15] the authors evaluated the performance of the soybean commodities' forecast prices and the forecast demand for perishable products using Wavelet Neural Networks (WNNs). The results were compared to the Backpropagation algorithm and the Extreme Learning Machine (ELM). This is the only work we found which addresses the ELM.

As can be observed, the linear models are still the most often used to perform coffee price forecasting. However, the accuracy of the results can be improved by using state-of-art approaches, especially nonlinear models. A relevant possibility is to resort the ELM, an ANN framework that is an efficient tool for solving prediction tasks. They are adequate for dealing with challenging forecasting problems since they present high approximation and generalization capabilities. The architecture is a simplified version of the traditional Multilayer Perceptron (MLP), but with a fast training process since the hidden neurons are not adjusted. Previous investigations present the feasibility of the proposal in similar contexts [10,12].

In this sense, we present a methodology to perform coffee price forecasting using Extreme Learning Machine (ELM) neural network as predictors [16,17] and preprocessing steps, like the treatment of data with a trend. In addition, we use the partial autocorrelation function (PACF) filter as a guide to determine the lags (inputs) [18].

We also applied the MLP [19] and linear models: Exponential Smoothing (ES), Autoregressive (AR), and Integrated Autoregressive and Moving Averages (ARIMA) models [20] to perform a comparative analysis of our proposal. To the best of our knowledge, an investigation containing this variety of approaches is unprecedented. Therefore, we intend to fill this gap.

The remainder of this work is organized as follows: In Section 2, we present the background and concepts related to forecasting methods (ES, AR, ARIMA, MLP, and ELM) used to predict coffee price. Also, we describe the methodology proposed in this research together with the methods for collecting, processing, and predicting data; Section 3 presents the computational results, performance comparison, and critical analysis; Section 4 considers the conclusions of this study and possible future work.

2. Materials and methods

In this section, the main concepts and modeling processes of the linear methods and neural networks addressed in this paper are described. We also present the metrics used to evaluate the models' performances, selection of inputs, and the complete methodology elaborated in this study to perform coffee price forecasting. Finally, we present the details of the collection of coffee price samples.

2.1. *Forecasting models*

Forecasting models are mathematical tools used to anticipate the behavior of a phenomenon to aid decision-making [20]. The most usual way to address this issue is to combine k previous values or lags $(y_{t-1}, y_{t-2}, \cdots, y_{t-k})$ of a time series to estimate the value $\hat{y}$of the phenomenon in time t. In other words, suppose that we would like to estimate the price of the coffee on October 10th. If we defined the use of the first three lags, the inputs $\mathbf{y}$ are October 9th, October 8th, and October 7th [18,20].

The process is summarized in founding the function F that maps the inputs in the desired output, as in Equation (1):

$$\hat{y}_t = F(y_{t-1}, y_{t-2}, \cdots, y_{t-k}) \tag{1}$$

Note that F can be determined using a linear or a nonlinear approach [19].

The coffee prices can be ordered in time considering the sampling such as hourly, daily, monthly, etc. These data form a time series, being adequate approaching forecasting models, which means that the lags are the coffee price values y_t used as input of the models. In this work, we address linear models (Exponential Smoothing, Autoregressive Model, and Integrated Autoregressive and Moving Average Model) and neural networks (Multilayer Perceptron and Extreme Learning Machines). The challenges are defining the most suitable model to determine F and selecting the best subset of lags (inputs). In this Section, we discuss our proposal to solve this task, which is based on the ELM.

2.1.1. *Exponential smoothing*

Exponential Smoothing (ES) techniques are widely used because of their ease of implementation, application, and adjustment and their achievement of efficient prediction results [21]. Among the ES models, there are three main proposals, which are considered as a basis for all other variations: the simple method, the Holt method, and the Holt-Winters method [22]. Each proposal was developed for different kinds of series, considering the presence of cyclicity, seasonality, and trend [23].

The simple ES method is used when there is no trend and seasonality in the time series. As a rule, the model demands only the past samples (lags) of the target variable and an approximation parameter with a value between 0 and 1, called the smoothing constant of the cycle (α) [21].

The Holt's Double Exponential Smoothing method can be used when the time series presents cyclicity and tendency but not seasonality. In addition to the parameter (α), this method uses the tendency smoothing parameter (β), with values ranging from 0 to 1 [22]. The technique is expressed according to Eq. (2). Eq. (3), and Eq. (4) represent the level estimate and the trend, respectively [20].

$$\widehat{Z}_{t+n} = L_t + nT_t \tag{2}$$

$$L_t = \alpha Z_t + (1-\alpha).(L_{t-1} + T_{t-1}) 0 \leq \alpha \leq 1 \tag{3}$$

$$T_t = \beta(L_t - L_{t-1}) + (1-\beta).T_{t-1} 0 \leq \beta \leq 1 \tag{4}$$

where $\widehat{Z}_{t+n}$ is the predicted value of the coffee for period $t+n$, (1 - α) is the exponential rate of the data weights, Z_t is the observed coffee price value for (time) period *t*, *n* represents the number of periods, and (1 - β) is the exponential rate of the weights of the data.

When the time series also presents a seasonal component, the Holt-Winters ES method is used. It is smoothed by a parameter (γ) [21]. Two types of seasonal effects (multiplicative and additive) are also considered in the calculation [21].

2.1.2. Autoregressive model

The first-order (*p*) Autoregressive (AR) model is a widely known method for time series modeling and forecasting. It requires the data series to present invariant variance, self-covariance, and mean over time [20], which means that it must be stationary [18]. The AR model is described by weighted past values and the random noise ε_t, as in Eq. (5):

$$\widehat{y}_t = \varphi_1 y_{t-1} + \cdots + \varphi_p y_{t-p} + \varepsilon_t \tag{5}$$

where $\widehat{y}_t$ is the predicted value in time *t*, y_{t-p} are the terms of the series (lags or the previous coffee price values), φ_p are the free coefficients of the model, and ε_t is the error intrinsic to the forecasting process.

Note that ε_t can be seen as a sequence of random and independent shocks generated from the process.

The definition of the free coefficient values of the model is made deterministically by the direct application of the Yule-Walker equations [20], which guarantees the global optimum values in the least mean quadratic error sense.

2.1.3. Autoregressive integrated moving average model

The Autoregressive Integrated Moving Average Model (ARIMA) is a combination of the Autoregressive (AR) and Moving Average (MA) models, added by differentiation (I) [24]. Described in [20], it aimed to capture the behavior of the autocorrelation or the serial correlation between the data of a time series to carry out prediction. Unlike AR and MA, ARIMA is an alternative that can be applied to a non-stationary series that presents a tendency behavior, since it satisfies the condition of data differentiation, making the data approximately stationary [8].

Some steps must be followed to perform predictions with ARIMA [20]: first, it is necessary to transform the series into a stationary pattern, where the number of differentiations is defined by *d*; then, the values *p* and *q* (orders of the AR and the MA, respectively) must be found in parallel with some method of estimating the free coefficient values. It is possible used the partial autocorrelation function (PACF) for this task [24]. Once the parameters are defined, the predictions can be made. The ARIMA(*p*, *d*, *q*) is described in Eq. (6):

$$y_t = \varphi_0 + \varphi_1 y_{t-1} + \cdots + \varphi_p y_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t \tag{6}$$

where $\varphi_0$is a constant in the estimated model, $\varphi_1, \cdots, \varphi_p$ are the weights for the past values of y_t (previous coffee prices values), $\theta_1, \cdots, \theta_q$ are the parameters that make it possible to write the series as a function of past shocks, and $\varepsilon_t$are the error terms, similar to Eq. (5).

It should be noted that the terms φ_p weigh the inputs from the AR and θ_q from the MA model. Also, it is common to consider $\varphi_0 = 0$.This forecasting method has presented good results and has been used in prediction problems in different contexts [25].

2.1.4. Multilayer perceptron

The Multilayer Perceptron (MLP) neural network is considered one of the most versatile architectures in terms of its potential applications [26]. According to [27], MLP is capable of approximating any continuous, nonlinear, differentiable, and limited function. Therefore, MLP is considered a universal approximator.

An MLP is composed of an input layer, an output layer, and one or more intermediate layers. The layers are made up of artificial neurons. The intermediate layers receive the input signals from the first layer. Next, they map these signals in a nonlinear way to another space, which is defined according to the problem. The transformed signal enters the output layer, which provides the network response [28].

The network has no recurrence: that is, it presents a feed-forward condition. Disjoint layers are connected, although their neurons are incommunicable. The signal processing of the network is given by Eq. (7) [29]:

$$y = f_s\left(\sum_{k=0}^{K} W_{1k}^{0}\left(f\sum_{n=0}^{N} w_{kn}^{i} u_n + b_n\right)\right) \tag{7}$$

in which u_n represents each input of the network, b_n is the bias, f is the activation function of the hidden layer, f_s is the activation function of the output layer, and the output signal is *y*. Note that vector u_n contains the previous coffee prices values used to estimate the future value *y*.

In Eq. (7), w_{kn}^{i} are the weights of the intermediate layer, $k = 0, 1, \ldots, k$ indicates the neuron at 0, 1, …, *N* (each input index), and W_{1k}^{0} are connections of the output neurons.

The training of an MLP is the process of adjusting the synaptic weights. The backpropagation algorithm, most commonly used, operates through a set of data patterns inserted in the network. At each data input, the network analyzes its outputs and then compares the responses obtained with the desired responses to calculate the error value and thus adjust the weights [26]. The backpropagation method is widely used to solve this task [29].

Please cite this article as: C. Deina, M. H. do Amaral Prates, M. S. R. Alves et al., A methodology for coffee price forecasting based on extreme learning machines, Information Processing in Agriculture, https://doi.org/10.1016/j.inpa.2021.07.003

It is crucial to define the stopping point of the process. Usually, the minimum training error criterion is used to interrupt the process when either the expected error value or a maximum number of iterations is reached [18].

The choice of the set of weights is often made using the hold-out cross-validation technique, which helps with the definition of the topology. This technique uses a set of samples of the network at each iteration. The one with the lowest validation error has its weights adopted as those able to maximize the generalization capacity of the network [19].

2.1.5. *Extreme learning machines*

To improve efficiency and reduce training bottlenecks in feedforward neural networks, the Extreme Learning Machine (ELM) was developed [30]. Huang et al. introduced an architecture quite similar to the MLP, with feedforward interconnections but only one intermediate layer. Like MLPs, the ELMs are capable of approximating any continuous, nonlinear, differentiable, and limited function [31,32].

The main difference between such approaches lies in the training process [30]. While in the MLP, the weights of the hidden layer and the output layer are adjusted by nonlinear optimization methodologies, in the ELM, the hidden layer weights are randomly generated and maintained untuned. Therefore, just the weights of the output layer are calculated [33].

If new neurons are inserted into the hidden layer, the network potential is increased to the required precision. According to [12], the activations $\mathbf{x}^h$ of these neurons are given by Eq. (8), and Eq. (9) generates the output, which are partially expressed in matrix form:

$$\mathbf{x}^h = f^h\left(\mathbf{W}^h\mathbf{u} + \mathrm{b}\right) \tag{8}$$

$$\mathrm{y} = \mathbf{W}^{\mathrm{out}}\mathbf{x}^h \tag{9}$$

The vector $\mathbf{u} = [u_n, u_{n-1}, \cdots, u_{n-K-1}]^{\mathrm{T}}$ contains the input signal, $\mathbf{W}^{\mathrm{h}}$ is the matrix containing the weights of the intermediate layer, and b is the bias of each neuron. The function $f^h(\dot{s})$ specifies the activation functions of the hidden neurons, and $\mathbf{W}^{\mathrm{out}}$ is the matrix containing the weights of the output layer.

A common way to determine the weights of the output layer is to use the Moore-Penrose pseudo-inverse [18], which, by the calculation of the minimum norm (Eq. (10)), finds an approximation for this type of system of nonlinear equations:

$$\mathbf{W}^{\mathrm{out}} = \left(\mathbf{X}_h^{\mathrm{T}}\mathbf{X}_h\right)^{-1}\mathbf{X}_h^{\mathrm{T}}\mathbf{d} \tag{10}$$

where $\mathbf{d}$ is the vector with the desired outputs, and $\mathbf{X}_h$ is the matrix containing the hidden layer outputs, and $\mathbf{X}_h^{\mathrm{T}}$ is its transposed matrix of $\mathbf{X}_h$.

As the training process is simplified, the computational cost to adjust the network is also reduced, which is another advantage [33].

2.2. *Performance assessment*

Different performance assessments can be used to compare the prediction results of distinct models. Among the most used are Mean Square Error (MSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) [18].

The MSE measures the average of the quadratic differences between the estimated values and the actual values [24]. Its mathematical representation is described in Eq. (11):

$$MSE = \frac{1}{N_S}\sum_{n=1}^{N_S}(d_n - y_n)2 \tag{11}$$

in which N_S is the number of samples, d_n is the desired output for the n-th sample, and y_n is the response given by the models.

The MAE calculates the mean error between the predicted value and the observed value. Its calculation is expressed in Eq. (12):

$$MAE = \frac{1}{N_S}\sum_{n=1}^{N_S}|d_n - y_n| \tag{12}$$

Lastly, the MAPE measures the absolute mean percentage error for each time period minus the actual values divided by the actual values [31], which is described in Eq. (13):

$$MAPE = \frac{100}{N_S}\sum_{n=1}^{N_S}\left|\frac{d_n - y_n}{d_n}\right| \tag{13}$$

In Expressions (12) and (13) the variables are the same from Equation (11). For all performance measures, the closer to zero the result is, the better the forecast result [18].

2.3. *Forecasting methodology*

To achieve the objectives of this research, an investigation was performed by collecting raw data on prices for two specific types of coffee: Arabica and Robusta. Before the application of the predictors, it is necessary to transform the database into approximately stationary data [20,24]. The performance of nonlinear models can also be increased with this step [18].

After collection, the data are analyzed to identify their behavior, which may be composed of a trend, cyclicity, seasonality, and a random term (fluctuation). A way of detecting such components is to apply the Cox-Stuart [34,35] and the Friedman [36,37] tests. If the series presents a trend, the data must be normalized using differentiation (Eq. (14)):

$$y'_t = y_t - y_{t-1} \tag{14}$$

in which y'_t is the difference between the current value y_t of the series and the previous one y_{t-1}.

The next step is the selection of the inputs, i.e., the previous samples of the series that are combined to perform the output response of the forecasting method. For the linear approaches, this means that we need to define the order of the models. In this work, we use the Partial Autocorrelation Function (PACF) as a way to guide such selection.

Next, the parameters of the models are calculated, the forecasting is performed, and the seasonal or trend components are reinserted. The computational results were compared using MSE, MAE, and MAPE, described in Section 2.6. Additionally, a statistical test is applied to guarantee that the results are significantly distinct. The predictions by the ES, AR, and ARIMA were made with the aid of Minitab® software, while for MLP and ELM, we used Python®. For all models, the predictions were performed for P = 1, 3, 6, and 12 steps

Please cite this article as: C. Deina, M. H. do Amaral Prates, M. S. R. Alves et al., A methodology for coffee price forecasting based on extreme learning machines, Information Processing in Agriculture, https://doi.org/10.1016/j.inpa.2021.07.003

ahead, using the direct method [32]. The complete flowchart summarizing the steps followed in this investigation is shown in Fig. 1.

2.4. *Data analysis and processing*

The samples were collected from the database available at the website of the Center for Advanced Studies in Applied Economics at the Luiz de Queiroz College of Agronomy (ESALQ) - University of São Paulo (USP), Brazil (available at http://cepea.esalq.usp.br).

These data refer to the monthly average prices of the ESALQ Coffee Indicator (commodities and futures exchange). The samples correspond to a 60 kg bag traded in the spot market of the state of São Paulo, Brazil [38]. The samples are in Brazilian Reais (R$), because this is the currency of the country. We highlight that the proposal introduced in this work can be extended to any currency. Regardless of this detail, we could have converted to an international currency and subsequently applied the proposal presented in this work; however, we prefer not to do this due to the originality of the data source.

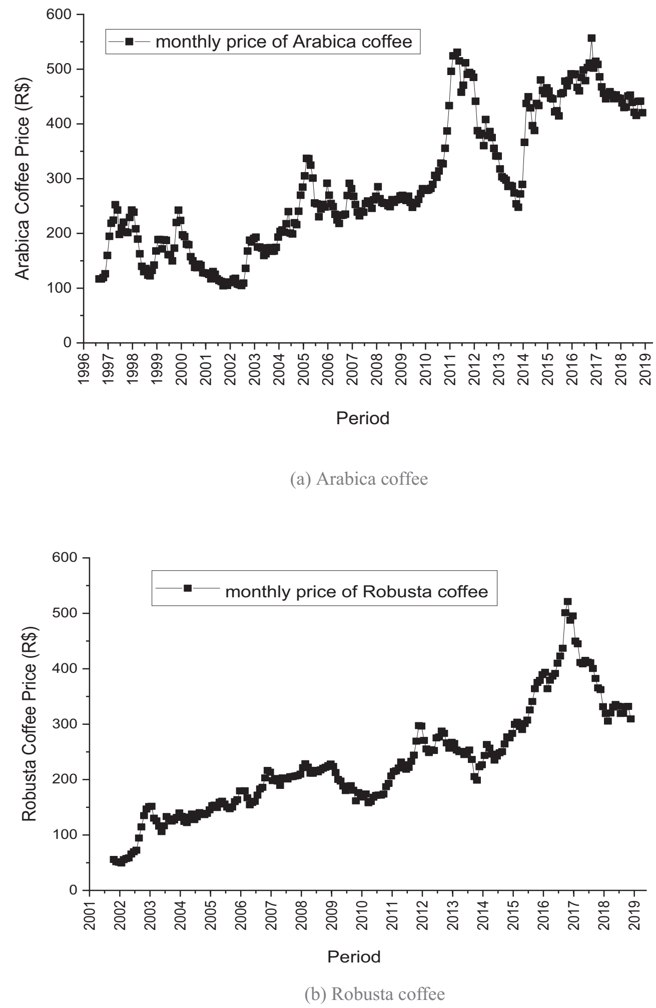
For Robusta coffee, monthly samples were collected uninterrupted from November 2001 to December 2018, totaling 206 prices. For Arabica, the data correspond to the period between September 1996 and December 2018, totaling 271 samples.

In 2018, the gross coffee billing was R$24.92 billion (US $8.5 billion) in Brazil. According to the Brazilian Ministry of Agriculture, Livestock, and Supply, 300,000 farmers are utilizing 2 million hectares [39].

The complete coffee chain currently generates 8 million direct and indirect jobs [39]. Beyond being the largest global coffee producer and exporter, Brazil is also an avid consumer. About 20.5 million 60 kg bags were sold in its internal market in 2018, making the country the second-largest consumer of this beverage in the world, behind only the United States [1].

These are vital pieces of information that expose the importance of the database, which reflects the global coffee price. Initially, the series were plotted on a graph to highlight the monthly price components for Arabica and Robusta coffees (Fig. 2).

Next, we applied the Cox-Stuart and Friedman tests to identify the statistical behavior of the series. It was found that both have a tendency of growth over time, cycles of variation in prices practiced, and an absence of seasonality. Thus, the differentiation method was used to make the series approximately stationary (Eq. (14)).



(a) Arabica coffee

(b) Robusta coffee

Fig. 2 – Monthly Price data for Arabica (a) and Robusta (b) coffee from September 1996 to December 2018 and November 2001 to 2018, respectively.

Then, the graph in Fig. 3 was elaborated to analyze the partial autocorrelation function (PACF) [20].

It can be seen that the Arabica coffee price series has a significant autocorrelation coefficient until the second delay, while the Robusta coffee series shows autocorrelation to the first and to the fourth delay, and then just the 11th and 12th delays. We use this information as an indication to determine the best inputs (lags) to perform the forecasting.

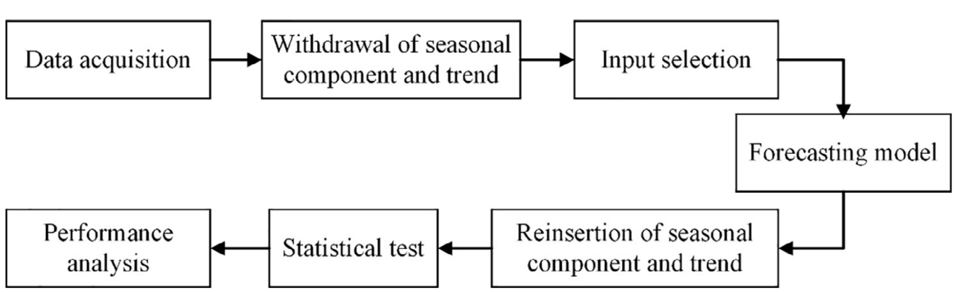


Fig. 1 – Complete flowchart for forecasting.

Please cite this article as: C. Deina, M. H. do Amaral Prates, M. S. R. Alves et al., A methodology for coffee price forecasting based on extreme learning machines, Information Processing in Agriculture, https://doi.org/10.1016/j.inpa.2021.07.003

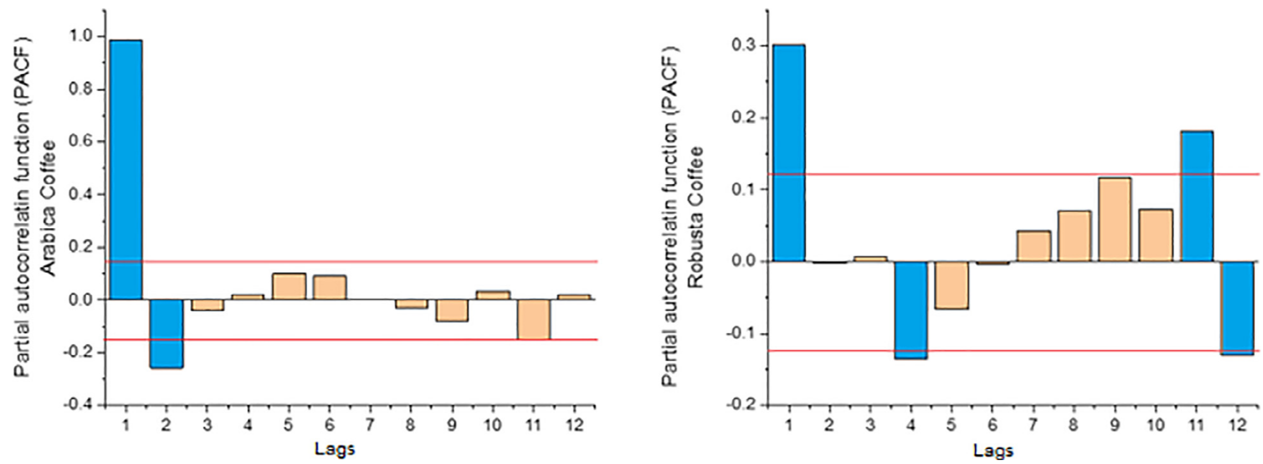


Fig. 3 – Graphics of the partial autocorrelation function (PACF) in function of different lags for the time series of (a) Arabica and (b) Robusta coffee prices. In blue are highligthed the most significant lags.

In general, the forecasting studies suggest addressing all significant lags (in blue) in the process, but we decide to analyze whether using all of them causes an increase in performance. We justify this decision, following some premises from relevant works that highlight that the variable selection of filter type (as the PACF) may not be the best choice depending on the predictor [40]. Hence, for each model, we performed an investigation on the best set of inputs based on the PACF.

3. Results and discussion

In this section, we present the results obtained with the analysis and treatment of the historical data of the prices of the Arabica and Robusta coffee. Subsequently, forecasts will be presented using the proposed methods. The data from 2017 to 2018 were used for testing, and the remainder was selected for training (adjustment of the models) to obtain predictions from the ES, AR, and ARIMA models. For MLP and ELM, the series up to 2015 was selected for training, the period from 2015 to 2016 was used for validation, and the samples from 2017 to 2018 were used for testing.

The choice of ES model depends only on the identification of the pattern in Fig. 2. From the analysis performed, the time series presented only a trend and cyclic and non-seasonal behavior, so it required the use of the dual Holt ES [24]. This model has as parameters the smoothing constants for the cycle and trend, α and β, respectively. However, there is no absolute method for defining these values [41]. Therefore, empirical parametric tests were performed until we reached the best results. The values chosen were $\alpha = 0.9$ and $\beta = 0.3$ for Arabica coffee and $\alpha = 0.9$ and $\beta = 0.6$ for Robusta coffee.

According to [41], large values of the constant α make the model react quickly to changes. Since the two series are quite variable over the months (Fig. 2), this may explain why the best value for α approached 1.

Regarding the selection of the input lags, the process was based on the PACF, as mentioned previously (Section 2.8), after the differentiation of the series. For Arabica, the order of the AR model was considered up to 2 delays, and for Robusta, it was considered up to 4. In the last scenario, the inclusion of the 11th and 12th delays did not bring an increase in performance. In both cases, the AR (1)—which considers just the first lag—reached the smallest errors.

Regarding the ARIMA, as the series required only one differentiation to make it stationary, the parameter d was defined as unitary [20]. The parameters p and q were chosen using the same procedures that were used for the AR, but in this case, the MA coefficients were defined for up to 4 inputs. The ARIMA (1,1,1) was chosen for Arabica, and the ARIMA (3,1,1) was chosen for Robusta coffee. Note that the maximum likelihood estimator was used for the adjustment of the model [24]. For data prediction using both neural networks, the number of neurons in the hidden layer was defined empirically, starting with 3 neurons and increasing by 5 up to 50 neurons. For each variation, the ANNs were executed thirty times [25]. For Arabica coffee, the ANNs used delays 1 and 2, while for Robusta they used from 1 to 4 lags, following the same premises of the AR model.

After the definition of the parameters of the models, the prices series for 1, 3, 6, and 12 steps ahead were predicted for the 24 months of the test set. Table 1 shows the computational results obtained from the test sets regarding the Mean Square Error (MSE), Mean Absolute Error (MAE), and Mean Absolute Percent Error (MAPE), while Table 2 presents a ranking of the performances, considering the forecasting models addressed. The lowest values per horizon are highlighted in bold, with the values found by neural networks referring to the best of the 30 simulations. Also, we present the number of neurons (NN) in the hidden layer of the ELM and MLP. The Diebold-Mariano test was applied to verify if the results were statistically different [36]. The p-values found were lower than 0.05, which ensures that changing the predictor leads to distinct results. Such p-values are shown in Table A1 in Appendix A. All tests have been executed with a confidence level of 95% [36].

Initially, it is important to highlight that in most cases, the prediction capability of the models decreases when the

Table 1 – MSE, MAE, and MAPE for all models and 1, 3, 6, and 12 steps ahead and the number of neurons used for each ANN.

		Arabica coffee				Robusta coffee			
	Model	P = 1	P = 3	P = 6	P = 12	P = 1	P = 3	P = 6	P = 12
MSE	ES	181.5528	594.3308	1227.3680	934.6249	274.7953	792.1526	1257.2049	2584.5590
	AR	172.8157	180.7207	186.6476	185.0037	288.6328	291.5460	295.7658	284.4327
	ARIMA	172.6134	179.4489	184.3301	181.3877	364.3127	352.0832	318.2893	321.4210
	MLP	155.3316	154.5009	153.4117	153.6183	**210.7712**	275.1311	**212.4111**	**232.6209**
	ELM	**125.5083**	**146.9868**	**146. 3150**	**144.2918**	233.2394	**235.2154**	220. 4867	245. 4732
MAE	ES	10.0713	19.7866	29.0799	23.9554	13.1242	22.1870	29.5507	45.5061
	AR	10.9242	10.8573	11.0704	10.9295	12.4184	13.3067	13.2874	13.0351
	ARIMA	10.8704	10.8307	11.0279	10.8317	13.4261	13.8219	13.4104	13.3649
	MLP	10.0556	10.1226	10.0750	10.0834	11.5682	12.1919	**11.1166**	**11.3502**
	ELM	**8.8055**	**9. 5835**	**9. 5418**	**9. 8699**	**11.2386**	**11.5442**	11. 4330	11. 4330
MAPE	ES	2.2444	4.4579	6.5349	5.3122	3.5923	6.3239	8.1818	12.4504
	AR	2.4069	2.4046	2.4492	2.4182	3.3445	3.6327	3.6125	3.5356
	ARIMA	2.3979	2.4010	2.4421	2.3985	3.5732	3.7576	3.6344	3.6142
	MLP	2.2168	2.2481	2.2375	2.2394	3.2322	3.3413	**3.0388**	**3.1180**
	ELM	**1. 9362**	**2.1544**	**2.1660**	**2.1652**	**3.2085**	**3. 1494**	3.1724	3.1943
NN	MLP	15	10	5	5	30	10	30	20
	ELM	25	10	10	5	5	25	30	15

Table 2 – Ranking of performances for all models and 1, 3, 6, and 12 steps ahead, with 1 being the best performance and 5 being the worst.

		Arabica Coffee				Robusta Coffee			
	Model	*P* = 1	*P* = 3	*P* = 6	*P* = 12	*P* = 1	*P* = 3	*P* = 6	*P* = 12
MSE	ES	5	5	5	5	3	5	5	5
	AR	4	4	4	4	4	3	3	3
	ARIMA	3	3	3	3	5	4	4	4
	MLP	2	2	2	2	1	2	**1**	**1**
	ELM	**1**	**1**	**1**	**1**	2	**1**	2	2
MAE	ES	3	5	5	5	4	5	5	5
	AR	5	4	4	4	3	3	3	3
	ARIMA	4	3	3	3	5	4	4	4
	MLP	2	2	2	2	2	2	**1**	**1**
	ELM	**1**	**1**	**1**	**1**	**1**	**1**	2	2
MAPE	ES	3	5	5	5	5	5	5	5
	AR	5	4	4	4	3	3	3	3
	ARIMA	4	3	3	3	4	4	4	4
	MLP	2	2	2	2	2	2	**1**	**1**
	ELM	**1**	**1**	**1**	**1**	**1**	**1**	2	2

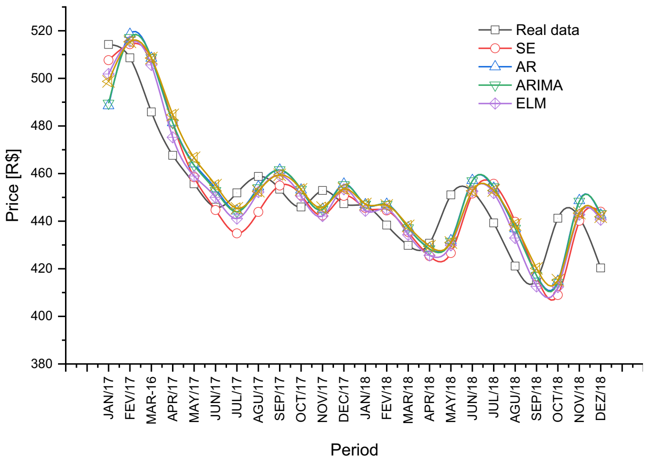
horizon grows. This behavior can be explained because the correlation tends to be high in nearby samples, showing an expected temporal dependency [43,44].

Regarding the Arabica coffee price, it is clear that our proposal of using the ELM obtained better performances for all horizons and metrics, overcoming the traditional MLP. This is an important result, since the ELM is easier to implement and understand, and its computational cost is smaller than that of the MLP. In addition, as discussed in Section 1.1, the MLP is almost always used as the neural approach for coffee price forecasting.

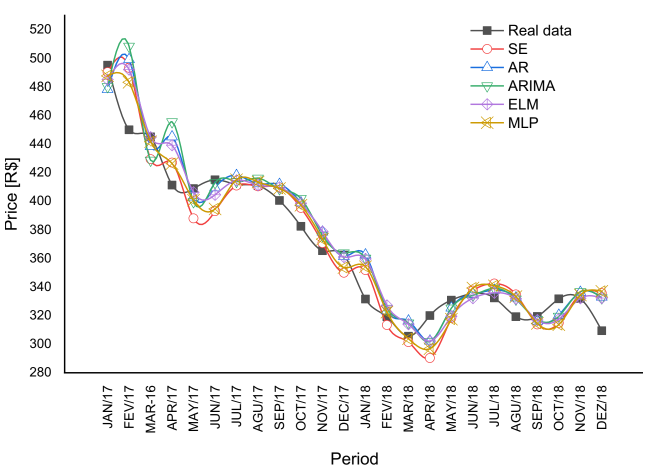
In regard to the Robusta type, there was almost a draw between ELM and MLP. The first stands out for *P* = 1 and 3 (except for MSE), while the MLP stands out for the other cases. Again, this is a major result, because the ELM was competitive once again. In terms of the overall performance of the models, the ELM overcame the other models in 17 out of 24 scenarios (71%) considering all metrics.

The linear models (ES, AR, and ARIMA) achieved worse prediction results for the two series analyzed when compared to the neural networks. In this case, the ARIMA performed better for the Arabica, while the AR performed better for the Robusta. The ES, in almost all cases, presented the worst errors.

Although the ARIMA presented more information to form the output response, sometimes the Maximum Likelihood

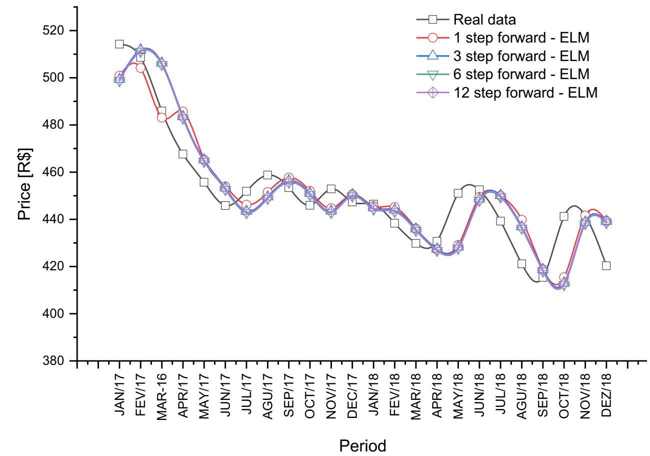


(a) Arabica coffee – 1 step forward

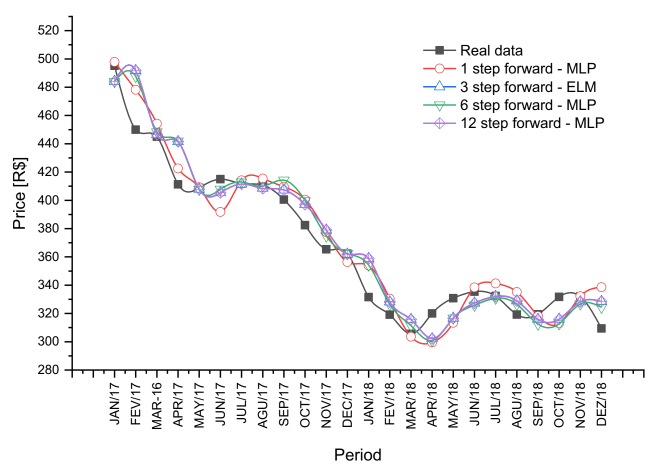


(b) Robusta coffee – 1 step forward

Fig. 4 – Graphics showing a comparison between real data and data forecasted for 1 step forward for the 24 months from 2017 to the end of 2018, using the ES, AR, ARIMA, ELM, and MLP models, to (a) Arabic coffee and (b) Robust coffee.



(a) Arabica coffee - the best performances of forecasting models



(b) Robusta coffee - the best performances of forecasting models

Fig. 5 – Graphics showing a comparison between real data and the best forecasting performances for 1, 3, 6, and 12 steps forward for the 24 months from 2017 to the end of 2018 among the methods ES, AR, ARIMA, ELM, and MLP, to (a) Arabica coffee and (b) Robusta coffee.

Estimator cannot achieve the best configuration for the parameters. On the other hand, the Yule-Walker equations guarantee the global optimum for the AR coefficients. Other optimizers could be used to adjust the ARIMA.

However, it is noteworthy that the linear models are simple in their implementation and can present adequate results, requiring less effort compared to the ANNs. Also, they are implemented and ready to use in many dedicated software applications, which present a friendly interface. This is one reason for their current, widespread use.

As the series were quite variable, the ANNs, being nonlinear techniques with high mapping capacity, presented the ability to adjust the predicted data to the real values. Moreover, their generalization capability is a significant reason for their good results [19].

To emphasize the gains in the use of ANNs, for Arabica, a decrease in the MSE of the ELM is observed to be $P = 1$ compared to the other models at a rate of 24% (MLP), 37% (ARIMA and AR), and 55% (ES). Analyzing the Robusta type, the MLP was better at a rate of 11% (ELM), 73% (ARIMA), 47% (AR), and 30% (ES). In relation to the other horizons, these amounts are even larger in most cases. Fig. 4 presents the performances of all models, considering one step ahead for the two data series analyzed. As can be observed, the neural

Table A1 – *p*-values found by the Diebold-Mariano test statistical tests.

	Arabica coffee (ELM)				Robusta coffee (MLP and ELM)			
Model	P = 1	P = 3	P = 6	P = 12	P = 1	P = 3	P = 6	P = 12
ES	0.0433	0.0285	0.0456	0.0037	0.0239	0.0446	2.68e-04	0.0459
AR	0.0051	0.0099	7.36e-05	1.00e-04	0.0259	1.98e-06	4.00e-07	7.10e-07
ARIMA	0.0047	0.0106	4.22e-05	5.06e-05	0.0470	9.26e-05	5.22e-06	1.07e-07
MLP	2.15e-04	1.89e-27	2.75e-30	2.67e-30	–	2.92e-20	–	–
ELM	–	–	–	–	0.0232	–	2.77e-04	1.22e-20

models were able to follow the temporal development of the series better than the linear models.

Note in Fig. 5 that the behavior of the output models for P = 1 is the one that best follows the actual data.

4. Conclusion

This work proposed a methodology to predict the prices of Arabica and Robusta coffee based on the use of Extreme Learning Machine (ELM) neural networks. The steps followed include the preprocessing of the series and a variable selection procedure, which used the Partial Autocorrelation Function as a guide.

The ELM is quite similar to the classic Multilayer Perceptron (MLP) neural networks, but it requires a lower computational effort because its intermediate layer is not adjusted, and its training is based on a deterministic method.

The performance comparison is realized by employing the MLP and the Linear Exponential Smoothing (ES), Autoregressive, and Autoregressive Integrated and Moving Average (ARIMA) linear models, which are the models most used for coffee price forecasting. For all approaches, the horizons of 1, 3, 6, and 12 steps ahead were considered.

The computational results showed that the methodology using the ELM was capable of outperforming all linear procedures, and it was better than the MLP in 71% of the cases. Linear models present simplicity in their application, but as the data analyzed were quite variable, the ANNs showed more accurate prediction results for both datasets.

Future works can be developed using other neural network architectures, ensembles, and series from other countries. Optimization metaheuristics can be used as variable selection methods or to tune the linear models [46,47]. For instance, linear programming and multi-criteria approaches can be important integrated tools to support decision-making and manage resources [42,45].

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to thank the National Council for Scientific and Technological Development (CNPq - Brazil), processes number 405580/2018-5 and 315298/2020-0, and the Araucária Foundation, process number 51497, for their financial support.

Appendix A

Table A1 presents the p-value found by the Diebold-Mariano test. For each case, we compared the best predictor with the others.

REFERENCES

[1] International Coffee Organization. Coffee Market Report. URL: http://www.ico.org. 2020.

[2] Novanda RR, Sumartono E, Asriani PS, Yuliarti E, Sukiyono K, Priyono BS, et al. A comparison of various forecasting techniques for coffee prices. J Phys Conf Ser 2018;1114:012119. https://doi.org/10.1088/1742-6596/1114/1/012119.

[3] Consuegra M, Garcia-Verdugo J. Measuring the functional efficiency ofagricultural futures markets. Aust J Agric Resour Econ 2017;61(2):232–46.

[4] Cuaresma CJ, Hlouskova J, Obersteiner M. Fundamentals, speculation or macroeconomic conditions? Modeling and forecasting Arabica coffee prices. European Review of Agricultural Economics 2018;45:583–615.

[5] Bacci LA, Mello LG, Incerti T, Paiva AP, Balestrassi PP. Optimization of combined time series methods to forecast the demand for coffee in Brazil: A new approach using Normal Boundary Intersection coupled with mixture designs of experiments and rotated factor scores. Int J Prod Econ 2019;212:186–211.

[6] Milas C, Otero J, Panagiotidis T. Forecasting the spot spices of various coffee types using linear and non-linear error correction models. International Journal of Finance and Economics 2004;9:227–88.

[7] Naveena K. Statistical modelling for forecasting of price and export of Indian coffee Doctoral dissertation. India: Institute of Agricultural Sciences, Banaras Hindu University; 2015.

[8] Naveena K, Singh S, Singh A. Hybrid time series modelling for forecasting the price of washed coffee (Arabica Plantation Coffee) in India. International Journal of Agriculture Sciences 2017;9(10):4004–7.

[9] Naveena K, Singh Subedar, Rathod Santosha, Singh Abhishek. Hybrid ARIMA-ANN modelling for forecasting the price of Robusta coffee in India. International Journal of Current Microbiology and Applied Sciences 2017;6(7):1721–6.

[10] Ayyanathan N, Kannammal A. Share price time series forecasting for effective supply chain information exchange. International Journal of Logistics Systems and Management 2014;18(1):139. https://doi.org/10.1504/IJLSM.2014.062125.

[11] Ayyanathan N, Kannammal A. Combined forecasting and cognitive Decision Support System for Indian green coffee supply chain predictive analytics. In: In: Proceedings of the IEEE International Conference on Cognitive Computing and Information Processing (CCIP). p. 1–6.
[12] Fousekis P. Price co-movement and the hedger's value-at-risk in the futures markets for coffee. Agricultural Economics Review 2017;18:35–47.
[13] Yashavanth BS, Singh KN, Paul A, Paul RK. Forecasting prices of coffee seeds using Vector Autoregressive Time Series Model. Indian J Agric Sci 2017;87:754–8.
[14] Pradkhan E. Referential activity in agricultural futures markets: evidence from quantile regressions. Aust J Agric Resour Econ 2017;61:610–25.
[15] Puchalsky Weslly, Ribeiro Gabriel Trierweiler, da Veiga Claudimar Pereira, Freire Roberto Zanetti, Santos Coelho Leandro dos. Agribusiness time series forecasting using Wavelet neural networks and metaheuristic optimization: An analysis of the soybean sack price and perishable products demand. Int J Prod Econ 2018;203:174–89.
[16] Suchithra MS, Pai Maya L. Improving the prediction accuracy of soil nutrient classification by optimizing extreme learning machine parameters. Information Processing in Agriculture 2020;7(1):72–82.
[17] Huang Guang-Bin, Zhu Qin-Yu, Siew Chee-Kheong. Extreme Learning Machine: Theory and applications. Neurocomputing 2006;70(1-3):489–501.
[18] Siqueira Hugo, Boccato Levy, Luna Ivette, Attux Romis, Lyra Christiano. Performance analysis of unorganized machines in streamflow forecasting of Brazilian plants. Appl Soft Comput 2018;68:494–506.
[19] Haykin S. Neural Networks and Learning Machines. New York-USA: Prentice Hall; 2008.
[20] Box GEP, Jenkins GM, Reinsel GC, Ljung GM. Times series analysis: forecasting and control. New Jersey-USA: John Wiley & Sons; 2015.
[21] Taylor James W. Exponential Smoothing with a damped multiplicative trend. Int J Forecast 2003;19(4):715–25.
[22] Gardner Jr ES, McKenzie E. Seasonal Exponential Smoothing with damped trends. Manage Sci 1989;35(3):372–6.
[23] Winters Peter R. Forecasting sales by exponentially weighted moving averages. Manage Sci 1960;6(3):324–42.
[24] Hyndman RJ, Athanasopoulos G. Forecasting: principles and practice. Melbourne, Australia: Otexts; 2018.
[25] Siqueira H, Luna I, Alves TA, Tadano YS. The direct connection between box & Jenkins methodology and adaptive filtering theory. Mathematics in Engineering, Science and Aerospace 2019;10:27–40.
[26] Haykin S. Neural networks: A comprehensive foundation. Upper Saddle River-USA: Prentice-Hall; 1999.
[27] Castro LN. Fundamentals of natural computing: Basic concepts, algorithms, and applications. Boca Raton-USA: Chapman & Hall; 2006.
[28] Li Chen, Li Zhenbo, Wu Jing, Zhu Ling, Yue Jun. A hybrid model for dissolved oxygen prediction in aquaculture based on multi-scale features. Information Processing in Agriculture 2018;5(1):11–20.
[29] Samadianfard Saeed, Ghorbani Mohammad Ali, Mohammadi Babak. Forecasting soil temperature at multiple-depth with a hybrid artificial neural network model coupled-hybrid firefly optimizer algorithm. Information Processing in Agriculture 2018;5(4):465–76.
[30] Huang GB, Zhu QY, Siew CK. Extreme Learning Machine: A new learning scheme of feedforward neural networks. In: In: Proceedings of the international joint conference on neural networks. p. 985–90.
[31] Tadano YS, Siqueira, HV, Alves TA. Unorganized machines to predict hospital admissions for respiratory diseases. In: Annals of the 2016 IEEE Latin American Conference on Computational Intelligence (LACCI), Cartagena, Columbia, 2016: p. 1.
[32] Siqueira H, Luna I. Performance comparison of feedforward neural networks applied to stream flow series forecasting. Mathematics in Engineering. Science and Aerospace 2019;10:41–53.
[33] Yu Huihui, Yang Ling, Li Daoliang, Chen Yingyi. A hybrid intelligent soft computing method for ammonia nitrogen prediction in aquaculture. Information Processing in Agriculture 2021;8(1):64–74. https://doi.org/10.1016/j.inpa.2020.04.002.
[34] Cox DR, Stuart A. Some quick tests for trend in location and dispersion. Biometrika 1955;42(1):80–95.
[35] Sprent P, Smeeton NC. Applied nonparametric statistical methods. London – UK: Chapman & Hall/CRC Press; 2016.
[36] Diebold FX, Mariano RS. Comparing predictive accuracy. Journal of Business and Economic Statistics 2002;20 (1):134–44. https://doi.org/10.1198/073500102753410444.
[37] Adhikari Ratnadip, Agrawal RK. A linear hybrid methodology for improving accuracy of time series forecasting. Neural Comput Appl 2014;25(2):269–81.
[38] Center for Advanced Studies in Applied Economics at the Luiz de Queiroz College of Agronomy (ESALQ) - University of São Paulo (USP) (Centro de Estudos Avançados em Economia Aplicada ESALQ/USP). Coffe Price. URL: http://cepea.esalq.usp.br. 2019.
[39] Ministry of Agriculture, Livestock and Supply, Brazil. Coffee production in Brazil. URL: https://www.gov.br/agricultura/pt-br. 2020.
[40] Guyon I, Elisseeff A. An introduction to variable and feature selection. Journal of Machine Learning Research 2003;3:1157–82.
[41] Montgomery D, Johnson L, Gardiner J. Forecasting and Time Series Analysis. New York: McGraw-Hill; 1990.
[42] Thesari SS, Trojan F, Batistus DR. A decision model for municipal resources management. Manag Decis 2019;57 (11):3015–34. https://doi.org/10.1108/MD-05-2017-0500.
[43] Siqueira Hugo, Boccato Levy, Attux Romis Ribeiro de Faissol, Ribeiro Attux Romis, Lyra Filho C. Echo State Networks in Seasonal Streamflow Series Prediction. Learning and Nonlinear Models 2012;10:181–91. https://doi.org/10.21528/lmln-vol10-no3-art5.
[44] Siqueira Hugo Valadares, Boccato Levy, Attux Romis Ribeiro de Faissol. Echo State Networks for Seasonal Streamflow Series Forecasting. Lecture Notes in Computer Science 2012;7435:226–36.
[45] Leme Murilo Oliveira, Trojan Flavio, Xavier Antonio Augusto de Paula, Antonio Carlos Francisco. Digital Energy Management for Houses and Small Industries Based on a Low-cost Hardware. IEEE Latin America Transactions 2017;14:4275–8. https://doi.org/10.1109/TLA.2016.7786305.
[46] Puchta Erickson, Lucas Ricardo, Ferreira Felipe, Siqueira Hugo, Kaster Mauricio, et al. Gaussian Adaptive PID Control Optimized via Genetic Algorithm Applied to a Step-down DC-DC Converter. In: 12th IEEE International Conference on Industry Applications (INDUSCON). p. 1–6. https://doi.org/10.1109/INDUSCON.2016.7874509.
[47] Santos Pedro, Macedo Mariana, Figueiredo Elliackin, Santana Jr. Clodomir Joaquim, Soares Fabiana, Siqueira Hugo, et al. Application of PSO-Based Clustering Algorithms on Educational Databases. IEEE Latin American Conference on Computational Intelligence (LACCI) 2018;1–6. https://doi.org/10.1109/LACCI.2017.8285690..

Please cite this article as: C. Deina, M. H. do Amaral Prates, M. S. R. Alves et al., A methodology for coffee price forecasting based on extreme learning machines, Information Processing in Agriculture, https://doi.org/10.1016/j.inpa.2021.07.003