

Article

Comparative Analysis of Linear Models and Artificial Neural Networks for Sugar Price Prediction

Tathiana M. Barchi ^{1,2,†}, João Lucas Ferreira dos Santos ^{2,†}, Priscilla Basseto ^{2,†}, Henrique Nazário Rocha ^{3,†}, Sergio L. Stevan Jr. ^{3,4,†}, Fernanda Cristina Correa ^{3,4,†}, Yslene Rocha Kachba ^{5,†}, and Hugo Valadares Siqueira ^{1,2,3,4,*}

- ¹ Graduate Program Computer Sciences (PPGCC), Federal University of Technology–Paraná (UTFPR), Dr. Washington Subtil Chueire St., 330, Jardim Carvalho, Ponta Grossa 84017-220, Brazil; tathianabarchi@alunos.utfpr.edu.br
 - ² Graduate Program in Industrial Engineering (PPGEP), Federal University of Technology–Parana, Ponta Grossa 84017-220, Brazil; joalucasferreira.epa@gmail.com (J.L.F.d.S.); pri_bass@hotmail.com (P.B.)
 - ³ Department of Electrical Engineering, Federal University of Technology–Parana, Ponta Grossa 84017-220, Brazil; nazario.utfpr@gmail.com (H.N.R.); sstevanjr@utfpr.edu.br (S.L.S.J.); fernandacorrea@utfpr.edu.br (F.C.C.)
 - ⁴ Graduate Program in Electrical Engineering (PPGEE), Federal University of Technology–Parana, Ponta Grossa 84017-220, Brazil
 - ⁵ Department of Industrial Engineering, Federal University of Technology–Parana, Ponta Grossa 84017-220, Brazil; yslener@utfpr.edu.br
- * Correspondence: hugosiqueira@utfpr.edu.br
† These authors contributed equally to this work.

Abstract: Sugar is an important commodity that is used beyond the food industry. It can be produced from sugarcane and sugar beet, depending on the region. Prices worldwide differ due to high volatility, making it difficult to estimate their forecast. Thus, the present work aims to predict the prices of kilograms of sugar from four databases: the European Union, the United States, Brazil, and the world. To achieve this, linear methods from the Box and Jenkins family were employed, together with classic and new approaches of artificial neural networks: the feedforward Multilayer Perceptron and extreme learning machines, and the recurrent proposals Elman Network, Jordan Network, and Echo State Networks considering two reservoir designs. As performance metrics, the MAE and MSE were addressed. The results indicated that the neural models were more accurate than linear ones. In addition, the MLP and the Elman networks stood out as the winners.

Keywords: artificial neural networks; linear models; sugar price; forecasting



Citation: Barchi, T.M.; dos Santos, J.L.F.; Basseto, P.; Rocha, H.N.; Stevan, S.L., Jr.; Correa, F.C.; Kachba, Y.R.; Siqueira, H.V. Comparative Analysis of Linear Models and Artificial Neural Networks for Sugar Price Prediction. *FinTech* **2024**, *3*, 216–235. <https://doi.org/10.3390/fintech3010013>

Academic Editor: Riza Demirer

Received: 30 November 2023

Revised: 20 February 2024

Accepted: 5 March 2024

Published: 12 March 2024



Copyright: © 2024 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Sugar is an essential commodity in world trade due to its extensive use in the food industry. Sugar derivatives are also used to produce biofuels, which are likely to grow in the next decade [1,2]. When used for food purposes, sugar is mainly produced from sugarcane, predominantly in tropical and subtropical regions. In temperate regions, the most addressed raw material is sugar beet [3,4].

Observing the sugar prices in different regions of the world, it is evident that the global sugar market is extremely fragmented. A high price volatility has been observed in recent years [5]. The volatile nature of the market is further intensified by the effects of various support measures that benefit the subsector [6]. It is one of the commodities most strongly protected by governments, as they seek to protect producers from low prices by implementing various policy instruments, such as border measures, minimum price levels, and subsidies [7]. This product may suffer influence from the intern markets due to strict regulations applied in countries like Russia, India, and China. Note that these countries are

in the top ranking of the world's largest producers. Other factors influence the price, such as import rate and production volume control [8].

As specified by Van der Eng [9], sugar is a crucial commodity for economic development. The average annual production is around 1700 million tons, which covers 24 million hectares of land around the world [10]. In addition, a 15% growth in agricultural production is expected for the next decade, which can be attributed to improvements in production and high production intensity, followed by technological innovation [11].

According to the United States Department of Agriculture (USDA), global consumption increased from 171,582 metric tons in 2019/20 to 177,795 metric tons in 2020/21 [12]. Similarly, the per capita sugar consumption is expected to rise due to urbanization and increased consumption of processed foods [11]. In the USA, the consumption of food commodities is also expected to grow, mainly due to the population increase [13].

In this sense, understanding the dynamics of sugar prices between the domestic and international markets is of paramount importance for the success of companies operating in this sector. It plays a significant role in determining the rate of inflation, wages, salaries and various policies in an economy. In the case of cash crops such as sugarcane, the level of production affects the raw material cost and the competitive advantages in the market [14].

The literature on modeling international sugar markets and projecting their prices tends to use general equilibrium models [6] or partial equilibrium models [15–17]. These are usually recursive approaches that provide annual market balances for production, consumption, trade, and world prices over a projection period. Thus, the models aim to assess the effect of specific policies on the international market and their ability to incorporate a wide range of policy variables.

Another approach related to sugar price analysis is based on time series techniques [8,18–21]. In this case, the policy simulation possibilities are relatively narrow, but allow us to use data at a much higher frequency level (e.g., daily) than partial or general equilibrium models. Time series analysis and forecasting is addressed in different fields of research, such as economics, finance, health, engineering, etc. [22], which is the focus of the current investigation.

Several models have been suggested in the literature to improve the accuracy of sugar price forecasting. Ribeiro and Oliveira [23] proposed a hybrid model composed of artificial neural network Kalman filters to predict the monthly price of sugar in Brazil and India. Melo, Milioni and Júnior [24] used the Mixture of Local Experts Models (MLEM) to predict the daily and monthly price of American domestic sugar. Esam [25] applied the Autoregressive Integrated Moving Average (ARIMA) model to estimate the annual Egyptian sugar consumption and production. Mehmood [26] also applied the ARIMA model to predict the annual production of the sugar crop in Pakistan. Vishawajith et al. investigated the growth or decrease in the sugar crop area in India using annual data and the ARIMA model [27].

The application, study, and development of forecasting models have been a theme very much discussed in the last few decades [28], and such discussion has mainly focused on linear models. The models from the Box and Jenkins family stood out as some of the most used. However, due to the growing use of neural-inspired approaches, there is still room for evaluating the performance of nonlinear approaches in many fields, such as commodities prices. There is an important research gap that deserves to be filled. In this sense, this paper aims to investigate the use of feedforward and recursive Artificial Neural Networks (ANNs) as well as linear models for the price of sugar in several countries. The models addressed are the Multilayer Perceptron (MLP), Extreme Learning Machines (ELM), Elman Network, Jordan Network, Echo State Networks-ESN, Autoregressive Model (AR), and Autoregressive Moving and Average model (ARMA) [3,29–31]. The databases addressed are related to four of the most important markets: the European Union, United States, Brazil and the World.

Commodity prices broadly influence general price levels, which are of interest to central databases, policymakers, businesses, and consumers, in such a way that decisions

depend on their expectations of future inflation. Thus, understanding the price dynamics between the national and international markets is of great importance for the strategic planning of the sugar and alcohol sector [9]. This study may effectively help society, researchers, and scientists working in this area [32].

This paper is organized as follows: Section 2 presents the details of the forecasting models; Section 3 discusses the database addressed, pre-processing stages and the performance metrics used; Section 4 shows the results found by the models and a critical analysis; finally, this paper is concluded in Section 5.

2. Materials and Methods

Time series forecasting is the process of predicting the future value of time series based on past observations when the univariate approach is addressed. It is widely used in organizational environments and presents deep statistical bases [28,33]. Time series forecasting can be classified into four groups: forecasting based on time series decomposition, smoothing-based techniques, regression-based techniques, and machine approaches [34]. This section describes the main concepts of the linear methods and ANNs addressed.

2.1. Autoregressive Model (AR)

The Autoregressive (AR) model belongs to the Box and Jenkins methodology, which assumes a linear relationship between the current and past values [28]. More specifically, the process involves a linear combination of the variable and their own lagged values, $x_{t-1} + \dots + x_{t-p}$, multiplied by a coefficient ϕ_i and with added noise of a_t , according to the Equation (1) [35]:

$$\hat{x}_t = \sum_{i=1}^p \phi_i x_{t-i} + a_t, \quad (1)$$

where \hat{x}_t is the value predicted at time t and p represents the order of the model.

The AR model presents a closed-form solution to determine the optimal coefficients in terms of the mean square error, referred to as the Yule–Walker equation. Also, the partial autocorrelation function (PACF) can be used to calculate the order of the model [36].

2.2. Autoregressive Moving and Average Model (ARMA)

Another commonly used linear method from the Box and Jenkins methodology is the Moving Averages (MA) model. Unlike Autoregressive (AR), white noise terms a_t are combined [37], weighted by the coefficients θ_i . However, those terms can be modeled as the errors of the previous predictions' ε_t .

In this sense, the Autoregressive and Moving Average models (ARMA) are a combination of the aforementioned models, as in Equation (2):

$$\hat{x}_t = \sum_{i=1}^p \phi_i x_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + a_t \quad (2)$$

in which p and q describe the orders of AR and MA, respectively, i.e., the realization of series prediction by the ARMA model uses p previous signals and q previous noise.

Unlike the AR model, calculating the coefficients requires the solution of non-linear equations due to the reinsertion of the error. Thus, maximum likelihood estimators are addressed, as this is the standard procedure of the literature [28].

2.3. Multilayer Perceptron (MLP)

The Multilayer Perceptron is a feedforward neural network widely used due to its capability to approximate any nonlinear, continuous, limited, and differentiable function, being a universal approximator [38]. It is suitable for prediction, clustering, and channel equalization, among other tasks [39–41].

An MLP presents at least three layers of artificial neurons: input, hidden, and output layers. Note that the number of hidden layers can vary. The neurons in a layer are fully connected with the next one, but neurons of the same layer do not communicate [42].

The training of an MLP is the process of adjusting the weights of the connections between neurons. The goal is to find the best set of weights based on some error metric, usually the mean square error. Applying the backpropagation algorithm to determine all the network weights is common. In this case, the gradient vector of the cost function formed by the network output and the desired signal is calculated, followed by an iterative adjustment using the steepest decent gradient [43].

2.4. Extreme Learning Machine (ELM)

The architecture of an Extreme Learning Machine (ELM) is quite similar to the MLP. The main difference between them is in the training process [44]: while in an MLP, the hidden layer and output layer weights are adjusted by non-linear optimization methodologies, in an ELM, the weights of the hidden layer are randomly generated and kept untuned. Therefore, only the output layer weights are calculated [45], reducing the computational effort required during training [46]. The ELM architecture is in Figure 1.

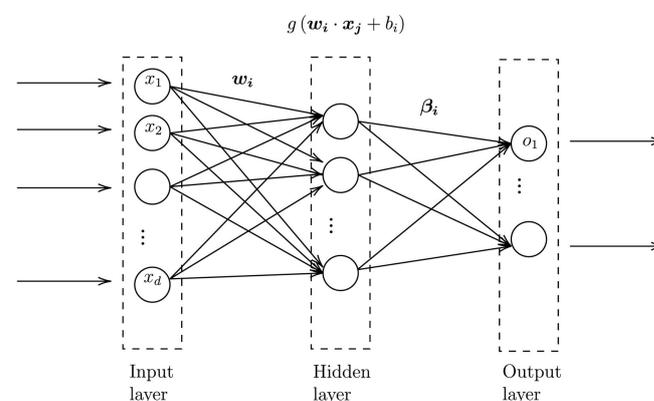


Figure 1. ELM architecture.

Using a constructive approach, the authors demonstrated that the model approximates any function by inserting new randomly generated neurons [47]. One can observe a direct relationship between ELM and MLP, the first “unorganized” version of the second, in which the middle layer is not subject to any adjustment [48].

In this sense, the ELM can approximate any nonlinear, continuous, limited, and differentiable mapping with arbitrary precision [47]. The necessary precision is achieved by inserting new neurons in the hidden layer, increasing the network’s approximation potential.

The network training consists of a linear regression that addresses the Moore–Penrose pseudo-inverse method to obtain the optimal values for the synaptic weights [44].

2.5. Elman and Jordan Networks

The Elman network presents a different characteristic from the previous ones: the existence of recurrent loops. Therefore, it is classified as a recursive architecture. The model presents a distinct input layer, which presents two parts: the first is the network’s input layer, and the second is responsible for storing the hidden layer outputs. This second part is called the context layer, and it contains adjustable synaptic weights. In short, the context layer stores the outputs produced by the intermediate layers, which are inserted as an input of the network [42].

Following similar premises, Jordan created the first recurrent neural network based on an MLP. This network was initially used to recognize time series, but nowadays, it is possible to find several applications in the literature. Unlike the Elman networks, the context units present synaptic weights and store the outputs produced by the output layer,

which are used as input together with the current external inputs. For both architectures, the truncated backpropagation through time is the method most frequently used to train the models [39,42].

2.6. Echo State Network (ESN)

The Echo State Network (ESN), proposed by Jaeger in 2001 [49], is a recurrent neural network proposal. ESNs generally comprise three layers: the input, the dynamic reservoir, and the output layer. The reservoir presents interconnected artificial neurons, generating a non-linear characteristic. The output layer is responsible for combining the outputs of the dynamic reservoir [39]. This proposal is as shown in Figure 2. The dynamic reservoir provides the network with an “internal memory” suitable for time-dependent problems.

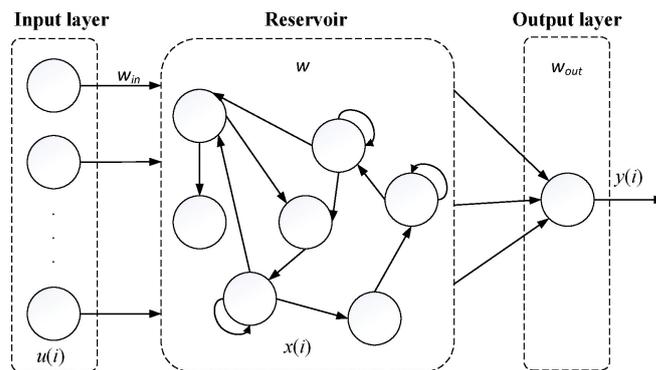


Figure 2. ESN architecture.

In Figure 2, W_{in} is the matrix containing the weights of the input layer connections, W the weights of the reservoir, and W_{out} the weights of the output layer. The last must be updated during the training process, while the others (W_{in} , W) are set randomly at the time of the creation of the network, without any changes [40,49].

The ESN presents similarities to the ELM as the training process is efficient due to the fast convergence and analytical solution for optimizing the output layer weights. Also, it is a universal approximation [39].

In this work, two reservoir designs were considered. The first, proposed by Jaeger [49], consists of a matrix of weights with three possible values randomly defined according to Equation (3).

$$W = \begin{cases} 0.4 & \text{with probability of 0.025,} \\ -0.4 & \text{with probability of 0.025,} \\ 0 & \text{with probability of 0.95.} \end{cases} \quad (3)$$

The second method considered was proposed by Ozturk et al. [50]. This proposal generates a reservoir rich in average entropy of the echo states. Furthermore, its eigenvalues present a uniform distribution in a unit circle, resulting in a canonical matrix according to Equation (4).

$$W = \begin{bmatrix} 0 & 0 & 0 & 0 & \dots & 0 & -r^N \\ 1 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 & 0 \end{bmatrix} \quad (4)$$

where r is the radial spectrum and N is the number of neurons in the reservoir.

In his pioneer work, Jaeger enunciated the “echo state propriety”, a mathematical proof of the existence of memory in an ESN [49]. If this condition is respected, the architecture

can stand untuned, and the output response will be stable. Ozturk et al. followed these premises [50].

To achieve the objectives of this research, an investigation was performed by collecting raw data on sugar prices from four databases: Brazil, the European Union, the United States, and the world. However, it is necessary to transform the database into approximately stationary before training the linear models. Note that the performance of nonlinear models can also be increased using this procedure [39].

The next step is the selection of the inputs, i.e., the previous samples (lags) of the series combined to perform the output response of the forecasting methods. For the linear approaches, this means defining the models' order. This work uses the Partial Autocorrelation Function (PACF) 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 aforementioned predictors were developed in Java. For all models (AR, ARMA, MLP, ELM, ELMAN, JORDAN, ESN-Jaeger, and ESN-Ozturk et al.), the predictions were performed for $P = 1, 3, 6,$ and 12 steps ahead, using the direct method [43,51].

3. Data Analysis and Processing

This section presents the details of the databases used in this work. The historical data from the European Union, the United States, and the World are available on The World Bank website [52]. The samples for Brazilian sugar prices were obtained at the Center for Advanced Studies in Applied Economics-CEPEA [53].

The database from Brazil (Figure 3) corresponds to monthly prices in Reais (R\$) (Brazilian currency) per kilogram and comprises the period from May 2003 to March 2020, totaling 203 samples. The European Union (Figure 4), the United States (Figure 5), and the World (Figure 6) databases correspond to monthly prices per kilogram in US dollars (US\$) and comprise the period from January 1960 to October 2019, totaling 718 samples. The series was collected until March 2020.

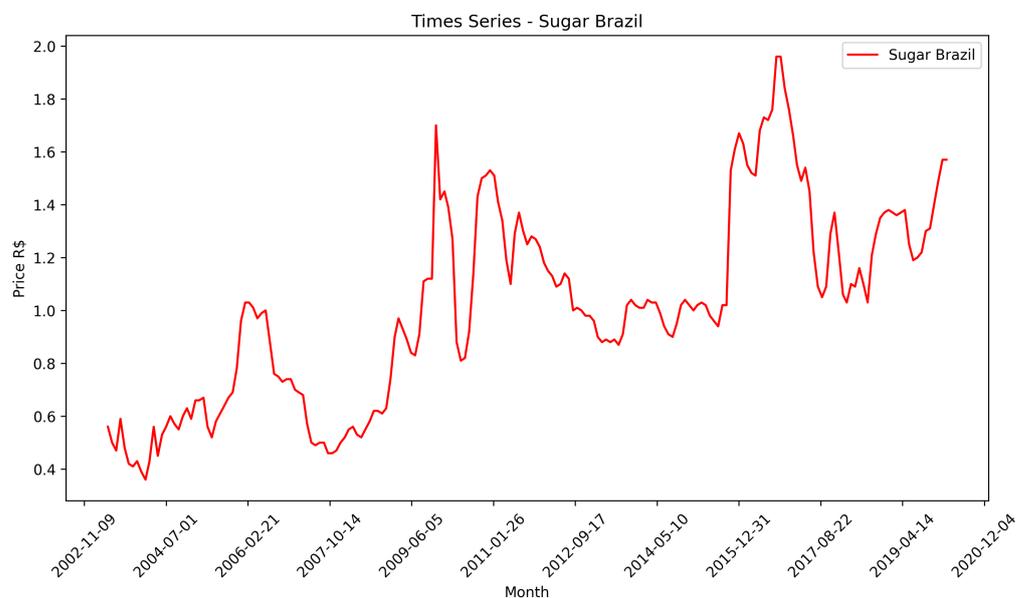


Figure 3. Sugar price data—Brazil.

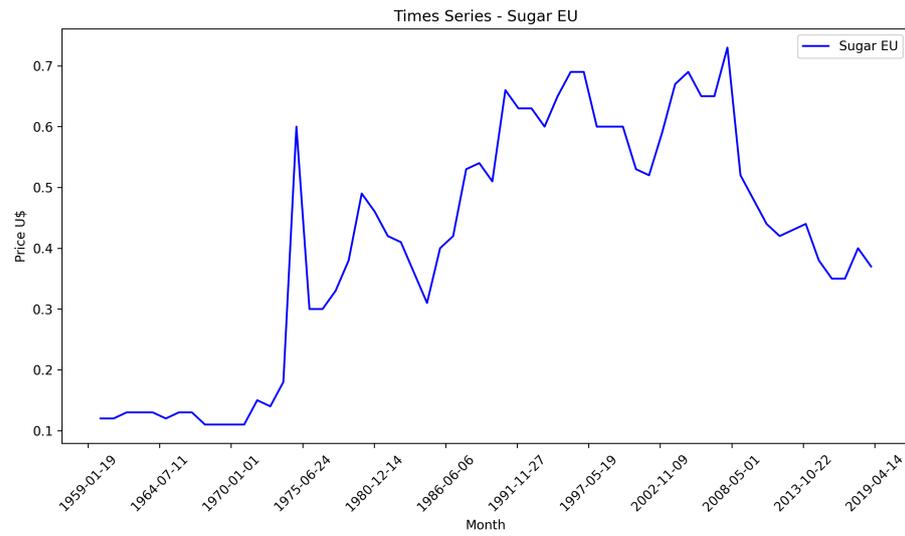


Figure 4. Sugar price data—EU.

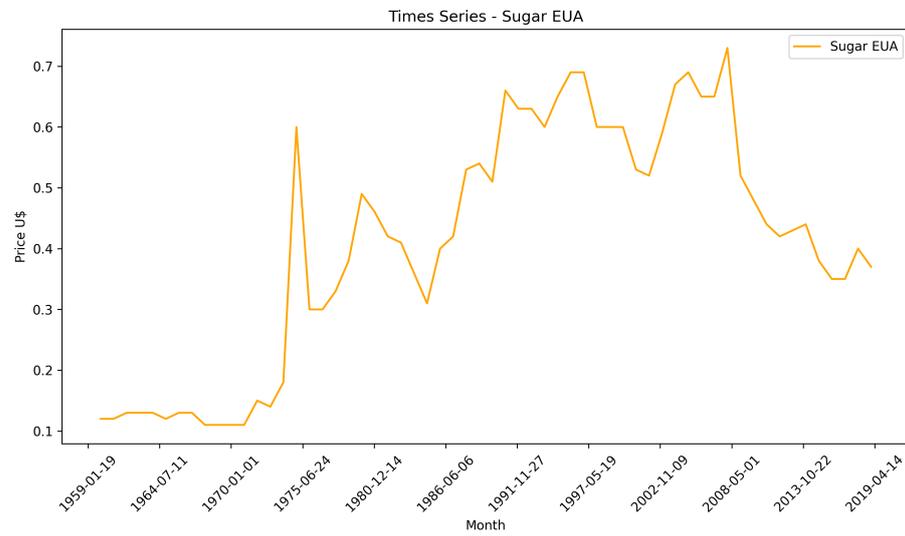


Figure 5. Sugar price data—USA.

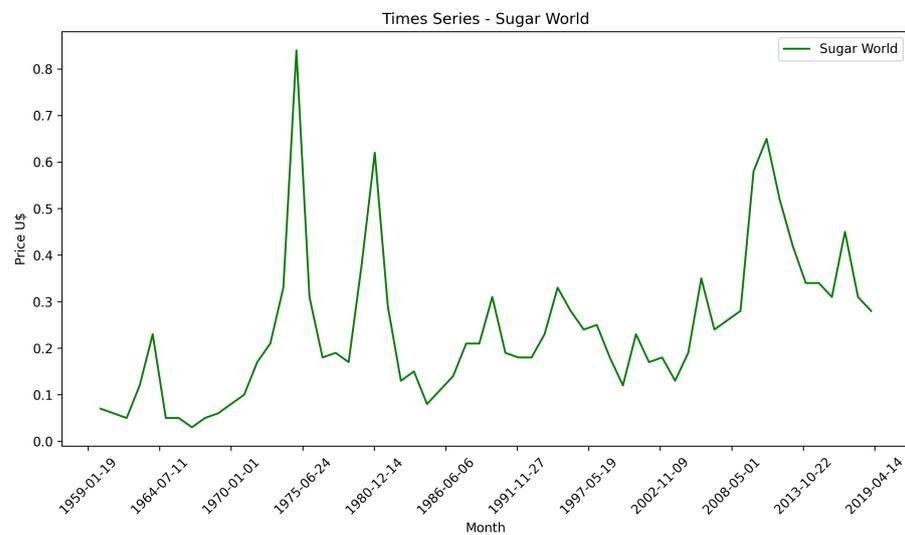


Figure 6. Sugar price data—world.

Figures 7–10 present the STL decomposition of each series into trend, seasonality and residual components [54].

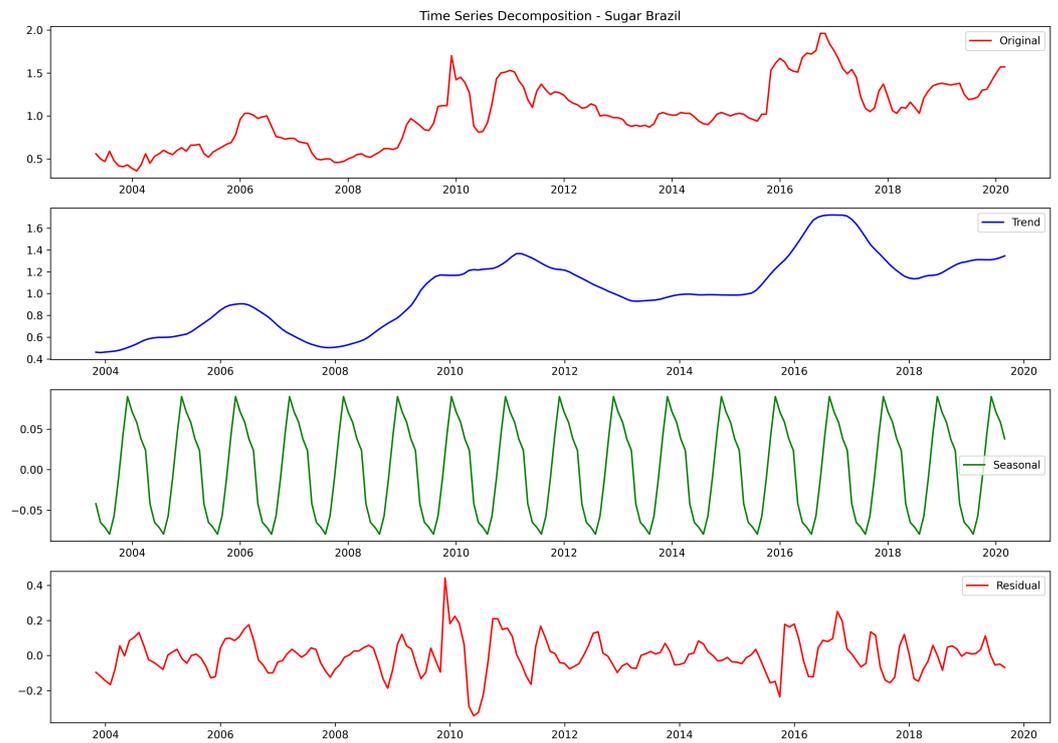


Figure 7. Decomposition—Brazil.

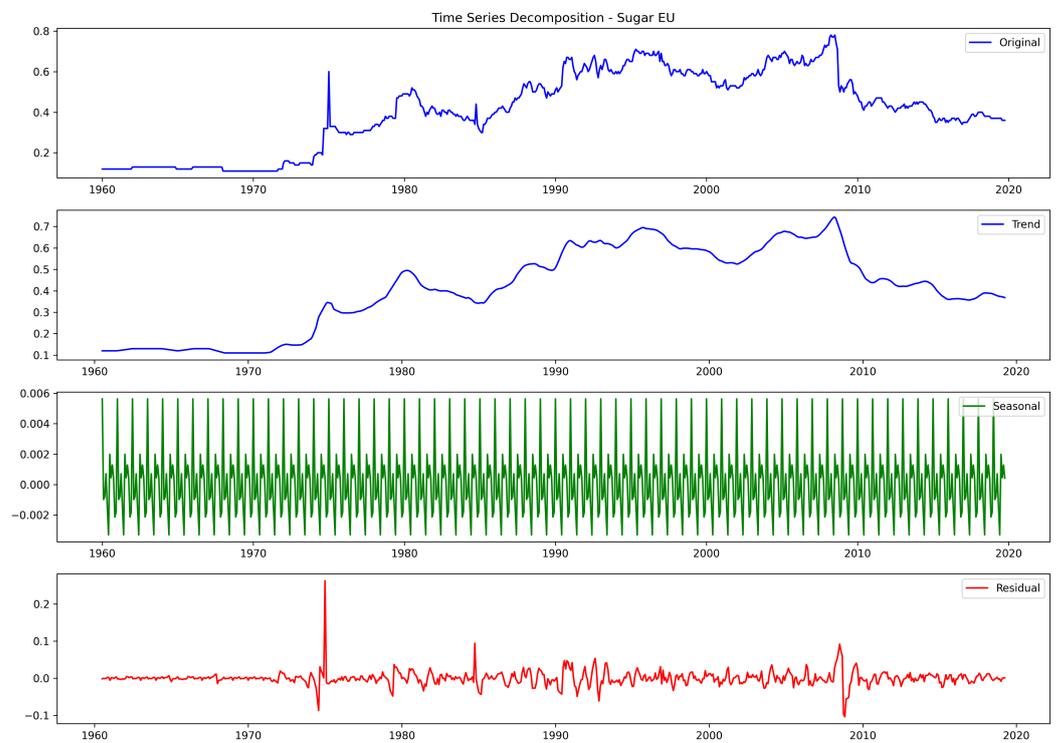


Figure 8. Decomposition—EU.



Figure 9. Decomposition—EUA.

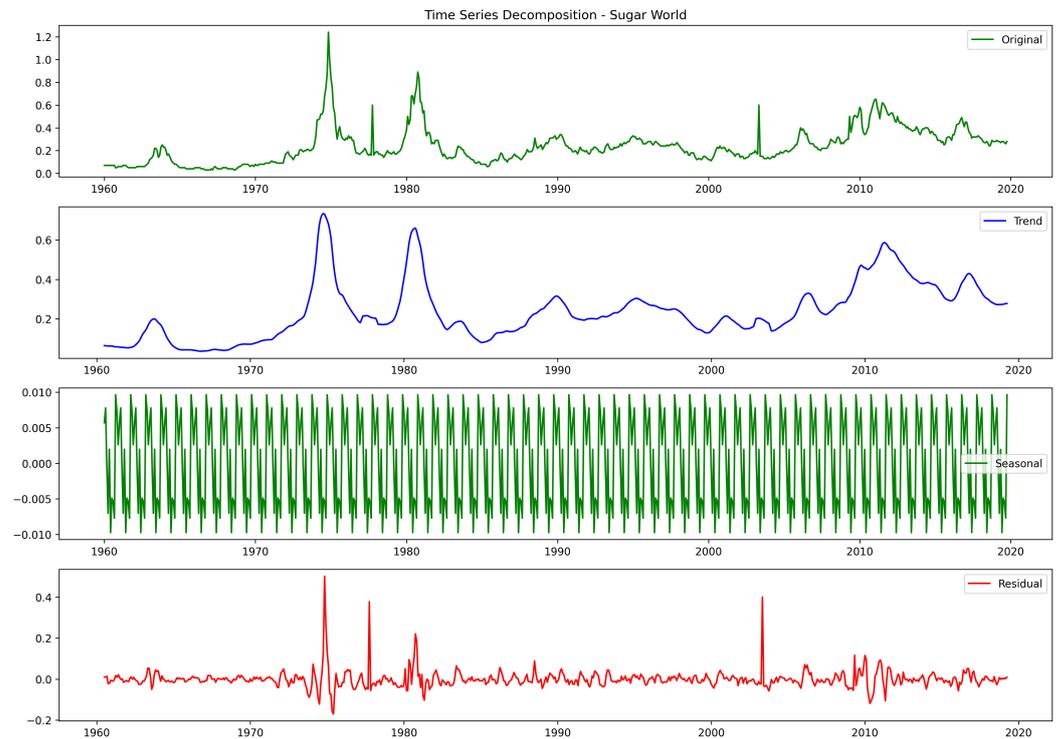


Figure 10. Decomposition—world.

In order to detect the presence of the trend component, the Kendall test was applied for each series. This component was detected only in the United States series. Thus, the differentiation method made the series approximately stationary [28].

Seasonality can be identified using the Friedman test [28]. It was found that all series were not stationary because they present seasonal components. In this sense, applying an adjustment that excludes the seasonal influences is required. After this process, the series

becomes approximately stationary, presenting zero mean and unitary standard deviation, which are mandatory for applying the linear methods. The seasonal adjustment function used in this work was the Z-Score, given by Equation (5):

$$z_{t,s} = \frac{x_{t,s} - \mu_s}{\sigma_s} \quad (5)$$

in which $z_{t,s}$ is the new adjusted price, $x_{t,s}$ represents the original sugar price in time t , μ_s is the mean, σ_s is the standard deviation, and $s = 1, \dots, 12$ is the seasonality index (in this case, the index of each month) [43].

Next, Figures 11 and 12 were expanded upon to analyze the autocorrelation function (ACF) and partial autocorrelation function (PACF) of each series, respectively. To achieve this, we considered up to 80 lags and the adjusted series.

The ACF of the E.U. and USA present a slow and constant decay, considering the decrease in the correlation of the samples. On the other hand, the ACF of the World and Brazil show an exponential decay. These formed functions present characteristics analogous to those of a combination of sine and damped exponentials, indicating that the generating process is associated with an autoregressive model [28].

Regarding the PACF, the E.U. series presents up to 6 significant lags, the highest value. The other series are limited to 4 lags [28].

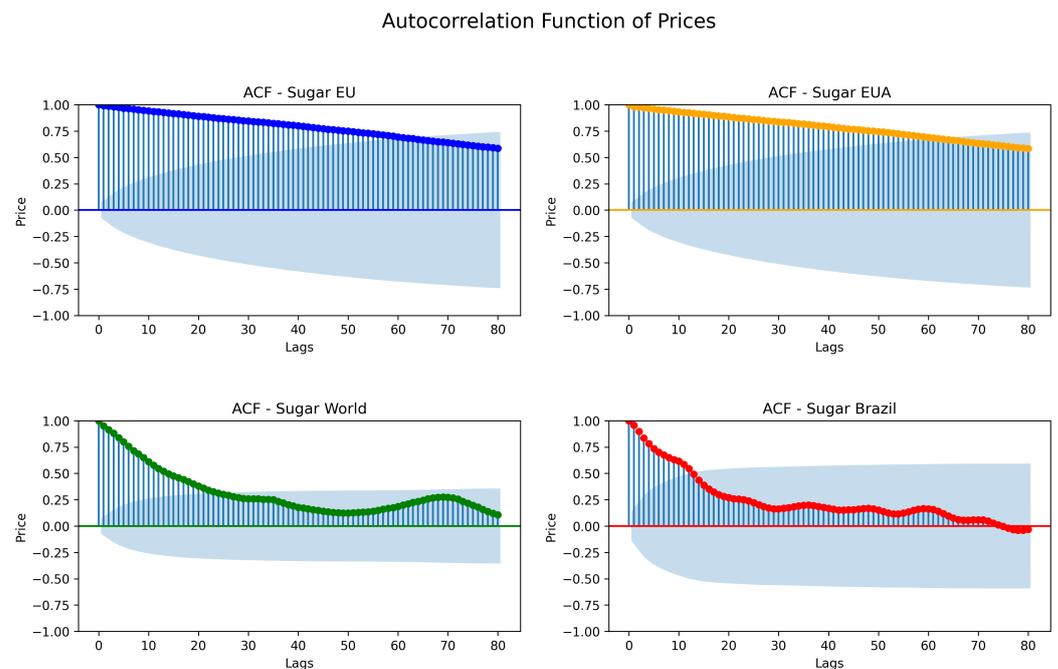


Figure 11. Autocorrelation Function.

To perform the experiments, a separation in the dataset was split into three groups—training, validation, and testing—in which the temporal order is respected. The first 66.67% were used for training, the following 17.01% for validation, and the last 16.32% for testing. Such separation is necessary for the application of Artificial Neural Networks. It is unnecessary for the AR and ARMA models to carry out validation, so the training and validation sets were used to adjust the models. This way, the test set presents the same size and samples for all experiments.

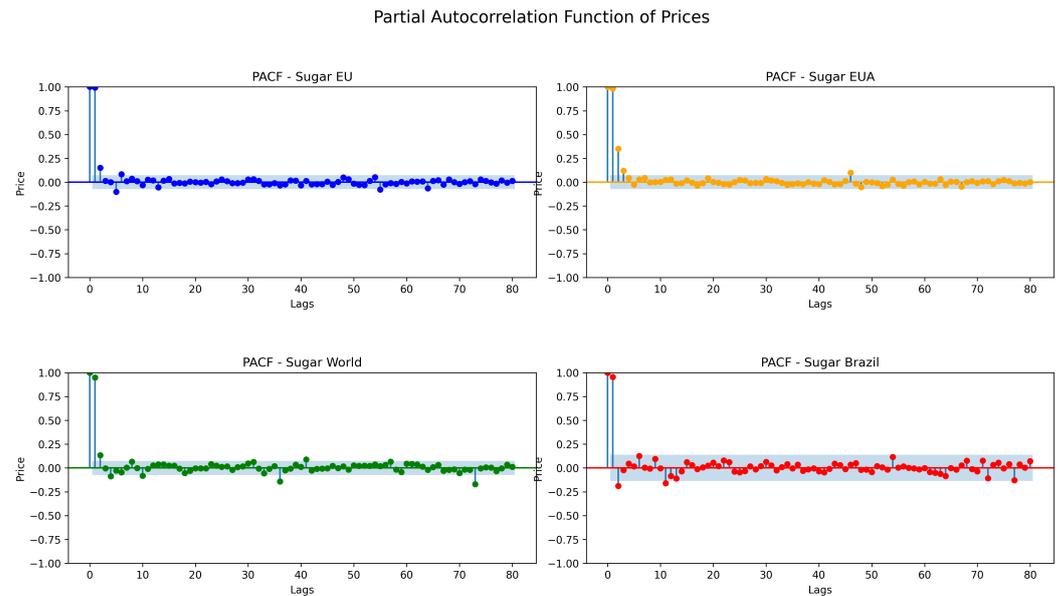


Figure 12. Partial Autocorrelation Function.

The wrapper method was used to select the most relevant inputs for the ANNs, which performs a systematic search for the best variables [41]. An exhaustive search was carried out for the AR and ARMA models, considering up to order 6. Table 1 presents the selected lags for the ANNs, the p order of the AR models, and the p and q orders of the ARMA models, considering the four series addressed. In Table 1, the inputs are problem-dependent, meaning that the best entries often differ when using the wrapper methodology.

Table 1. Input lags for each model according to the database.

Series	Models	
Brazil	AR	$p = 3$
	ARMA	$p = 1, q = 6$
	MLP	(1, 2)
	ELM	(1, 2, 3, 5)
	Elman	(1, 2)
	Jordan	(1, 2, 6)
	ESN-Jaeger	(1, 2, 3)
	ESN-Ozturk et al.	(1, 2, 4, 5)
European Union	AR	$p = 1$
	ARMA	$p = 1, q = 2$
	MLP	(1, 5)
	ELM	(1, 5)
	Elman	(3, 5)
	Jordan	(1)
	ESN-Jaeger	(1, 5)
	ESN-Ozturk et al.	(1, 3)
United States	AR	$p = 2$
	ARMA	$p = 1, q = 6$
	MLP	(1, 6, 5, 4)
	ELM	(1, 4, 5, 2)
	Elman	(3, 1, 4)
	Jordan	(1, 2, 3, 4, 5)
	ESN-Jaeger	(1, 5)
	ESN-Ozturk et al.	(1)

Table 1. Cont.

Series	Models	
World	AR	$p = 3$
	ARMA	$p = 2, q = 4$
	MLP	(1, 2)
	ELM	(1, 2)
	Elman	(1, 2)
	Jordan	(1, 2)
	ESN-Jaeger	(1, 2, 4)
ESN-Ozturk et al.	(1)	

Performance Assessment

To evaluate the results obtained some of the error metrics most frequently used in forecasting tasks are addressed: Mean Absolute Error (MAE), Mean Squared Error (MSE) [41,43]. They are given by the Equations (6) and (7):

$$MAE = \frac{1}{n} \sum_{i=1}^n |e_i| \quad (6)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n e_i^2 \quad (7)$$

in which $|e_i|$ represents the absolute error, that is, the difference between the predicted value and the real value for n samples.

4. Results and Discussion

This section presents and discusses the forecasting results achieved for Brazil, the European Union, the United States, and world sugar prices. The forecasting methods are shown in Section 2.

To determine the number of neurons for the unique hidden layer of each neural network, a grid search considering from 5 up to 50 neurons was performed for each architecture. Also, 30 independent rounds of simulation were executed for the neural models.

Tables 2–5 show the best result of the test set for 30 rounds, considering 1-, 3-, 6- and 12 steps-ahead forecasting, respectively. In these tables, NN denotes the number of neurons in the hidden layer of the best configuration. The values in bold represent the most minor errors for each database.

The Wilcoxon test was applied to the results regarding the MSE in the test set. The p -values achieved were close to zero (below 10^{-6}). Therefore, we can assume that changing the predictor implies a significant change in the results [29].

It is possible to notice that as the forecasting horizon increases, there is a tendency for errors to increase. This is partly due to the decreasing correlation between the desired inputs and outputs. Figure 11 reveals this behavior, as the lags are less correlated as the temporal distance among the samples increases. Also, the best predictor for MSE is often different for MAE. This behavior can occur since the used metrics penalize the difference between the desired response and the model's output differently. The MSE strongly penalizes the highest differences, since it uses the square operation. This was observed in previous studies on time series forecasting [40,41].

Table 2. Errors achieved by the forecasting models considering 1 step-ahead prediction.

	Error Metric	Brazil	E. Union	USA	World
AR	MSE	0.006619	0.000085	0.000507	0.000862
	MAE	0.062140	0.006841	0.016897	0.021225
ARMA	MSE	0.029678	0.000089	0.001591	0.002554
	MAE	0.131995	0.007073	0.029961	0.036826
MLP	NN	10	20	20	20
	MSE	0.007094	0.000119	0.000497	0.000838
	MAE	0.059427	0.008800	0.016724	0.020473
ELM	NN	10	15	15	5
	MSE	0.006787	0.000080	0.000505	0.000807
	MAE	0.062698	0.006924	0.017099	0.020540
ELMAN	NN	5	15	15	5
	MSE	0.007562	0.000303	0.001007	0.000907
	MAE	0.061830	0.013766	0.024728	0.021884
JORDAN	NN	10	20	20	5
	MSE	0.008757	0.000109	0.000525	0.000810
	MAE	0.074623	0.008644	0.017555	0.020608
ESN-Jaeger	NN	5	20	20	5
	MSE	0.006590	0.000083	0.000477	0.000786
	MAE	0.061112	0.007209	0.016723	0.020070
ESN-Ozturk et al.	NN	10	20	20	5
	MSE	0.007216	0.000088	0.000474	0.000808
	MAE	0.068246	0.007391	0.016793	0.020520

Table 3. Errors achieved by the forecasting models considering 3 steps-ahead prediction.

	Error Metric	Brazil	E. Union	USA	World
AR	MSE	0.039494	0.000312	0.001990	0.003883
	MAE	0.147834	0.013943	0.033760	0.044318
ARMA	MSE	0.061513	0.000322	0.003346	0.004903
	MAE	0.189204	0.013977	0.043025	0.051552
MLP	NN	10	20	10	20
	MSE	0.044594	0.000595	0.001825	0.003931
	MAE	0.158258	0.019043	0.030379	0.045695
ELM	NN	10	15	20	5
	MSE	0.049601	0.000287	0.002048	0.003379
	MAE	0.179047	0.013966	0.034950	0.042695
ELMAN	NN	5	5	5	5
	MSE	0.038662	0.000323	0.004625	0.018931
	MAE	0.146455	0.013696	0.059707	0.112465
JORDAN	NN	10	20	10	5
	MSE	0.089699	0.000519	0.002136	0.004182
	MAE	0.249985	0.018776	0.033911	0.048144
ESN-Jaeger	NN	5	20	10	5
	MSE	0.047391	0.000371	0.003045	0.005412
	MAE	0.175008	0.015139	0.039209	0.052192
ESN-Ozturk et al.	NN	10	5	10	5
	MSE	0.147545	0.000336	0.003698	0.004861
	MAE	0.30956	0.014135	0.044612	0.051995

Table 4. Errors achieved by the forecasting models considering 6 steps-ahead prediction.

	Error Metric	Brazil	E. Union	USA	World
AR	MSE	0.069570	0.000842	0.004466	0.007233
	MAE	0.203588	0.022235	0.048016	0.064934
ARMA	MSE	0.098970	0.000910	0.007033	0.007155
	MAE	0.247330	0.022679	0.061677	0.063863
MLP	NN	10	20	10	20
	MSE	0.070732	0.001899	0.003691	0.008879
	MAE	0.215856	0.032507	0.043801	0.076295
ELM	NN	10	15	20	5
	MSE	0.113170	0.000965	0.005450	0.006052
	MAE	0.254439	0.022597	0.055213	0.059288
ELMAN	NN	5	5	5	5
	MSE	0.061191	0.000714	0.015102	0.030949
	MAE	0.200087	0.021226	0.109164	0.146331
JORDAN	NN	10	20	10	5
	MSE	0.105431	0.002166	0.004839	0.007149
	MAE	0.269862	0.036479	0.053072	0.067726
ESN-Jaeger	NN	5	20	10	5
	MSE	0.104836	0.000940	0.005469	0.006602
	MAE	0.255797	0.023434	0.053360	0.058841
ESN-Ozturk et al.	NN	10	5	10	5
	MSE	0.090718	0.000631	0.007490	0.007044
	MAE	0.246907	0.020141	0.064420	0.064654

Table 5. Errors achieved by the forecasting models considering 12 steps-ahead prediction.

	Error Metric	Brazil	E. Union	USA	World
AR	MSE	0.110062	0.001794	0.012535	0.010816
	MAE	0.266367	0.031970	0.080744	0.087272
ARMA	MSE	0.206422	0.002089	0.018941	0.011183
	MAE	0.354384	0.033624	0.102893	0.085991
MLP	NN	10	20	10	20
	MSE	0.088511	0.005846	0.010390	0.019925
	MAE	0.250843	0.061302	0.079191	0.119357
ELM	NN	10	15	20	5
	MSE	0.143246	0.002846	0.020397	0.008954
	MAE	0.295927	0.037536	0.110122	0.078879
ELMAN	NN	5	5	5	5
	MSE	0.092892	0.001403	0.035525	0.042697
	MAE	0.247383	0.029694	0.170243	0.180500
JORDAN	NN	10	20	10	5
	MSE	0.120846	0.006725	0.017571	0.013186
	MAE	0.276443	0.065225	0.086754	0.097899
ESN-Jaeger	NN	5	20	10	5
	MSE	0.146639	0.002292	0.028207	0.009144
	MAE	0.300721	0.034410	0.100293	0.076992
ESN-Ozturk et al.	NN	10	5	10	5
	MSE	0.128009	0.001043	0.014985	0.009644
	MAE	0.291028	0.026485	0.094392	0.081244

By analyzing the values highlighted in bold, it is possible to observe that the linear models (specifically the AR) were the winners, only considering one step ahead of the European Union and the MAE. Furthermore, the ARMA and the Jordan network did not show the best results for any scenario.

Observing Table 2, it was noted that the Elman and Jordan networks did not present highlighted results. However, the other recurrent proposal, ESN, created by Jaeger and Ozturk et al., achieved the smallest MSE for three of four series and the best MAE for the world series. The other cases favored the MLP (best MAE for Brazil and the USA) and ELM (best MSE for the EU). In summary, considering the MSE, the most-used metric for forecasting purposes, the unorganized approaches (ELM and ESN) stood out. This is an interesting finding, since these methods are simple to implement and fast to adjust.

The results were different for three steps-ahead prediction. The ELM and Elman performed better, except for the USA, in which the MLP achieved the best errors.

Regarding the six- and twelve steps-ahead horizon, a pattern was observed. The Elman network, ESN-Ozturk et al., and MLP achieved the smallest errors for both metrics considering the Brazil, EU, and USA datasets, respectively. ELM and ESN-Jaeger were the best for the world dataset.

In order to show the general ranking of the models, Table 6 was created based on Borda Count Methodology [55], considering all forecasting horizons. In this case, “N. wins” means the number of wins each model achieved.

The linear models (AR and ARMA) achieved just one best result. This observation was expected since neural models are nonlinear techniques capable of approximating any dataset with a small error. The approach is more flexible than linear structures due to the presence of the nonlinear activation functions in the hidden layer, justifying their wide application in time series prediction [41,51,56]. However, it is essential to mention that the ANNs must be carefully adjusted to avoid overtraining. In this case, the model may “learn” the training samples, but it becomes unable to generalize or indicate a suitable response for unknown samples (test).

Table 6. Ranking of applied models considering all bases and different forecast horizons.

Models	1 Step		3 Steps		6 Steps		12 Steps		N. Wins
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	
AR		1							1
ARMA									0
MLP		2	1	1	1	1	1	1	8
ELM	1		2	1	1		1		6
ELMAN			1	2	1	1	1	1	7
JORDAN									0
ESN-Jaeger	2	1				1		1	5
ESN-Ozturk	1				1	1	1	1	5

Considering the general results of the ANNs, the Jordan network did not perform well. This is remarkable, since the Elman network, which presents a structure with a high similarity to the Jordan proposal, was the best for many cases. Indeed, the delayed output of the hidden layer seems to present more useful information in the composition of the formation of the response of the model than the output response.

The MLP was the architecture that achieved the best results, followed by the Elman network. As can be observed, the results are miscellaneous, but the Elman network was the best for multistep forecasting for Brazil, as well as the MLP for the USA.

The unorganized approaches performed well, too. We highlight the ESN-Jaeger for one step ahead and the ELM for three steps ahead. This is a good indication, since these methods are more accessible in adjusting the parameters, and there is no instability in the response due to a vanishing gradient, as in classic RNN approaches [42]. These results are

interesting because they corroborate the findings of previous studies that dealt with daily samples [3].

Finally, Figures 13–16 present the output of the best predictor according to the MSE, considering one step-ahead forecasting for the test set of Brazil, the European Union, the United States, and the world, respectively.

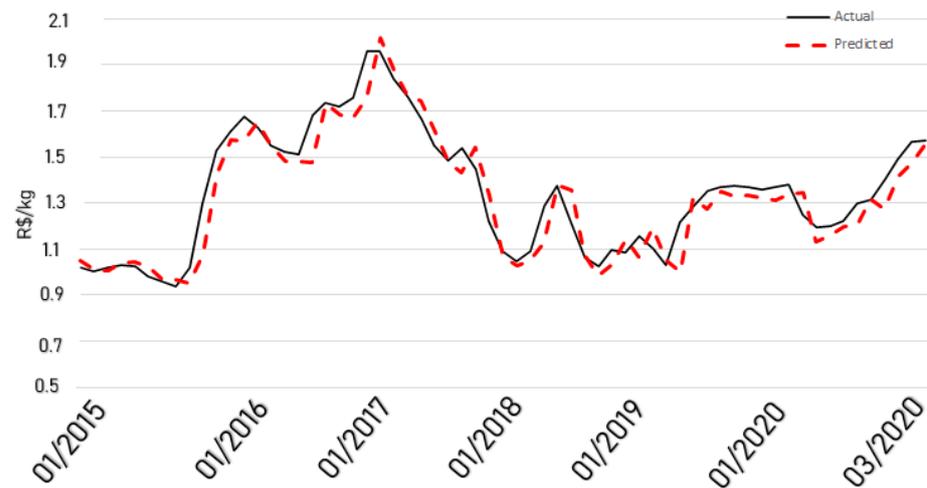


Figure 13. Brazil actual \times predicted.

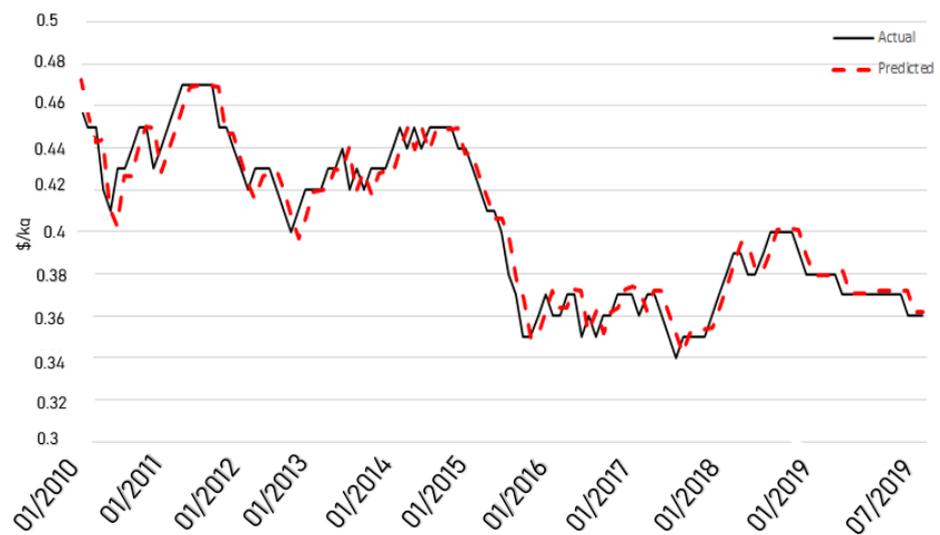


Figure 14. EU actual \times predicted.

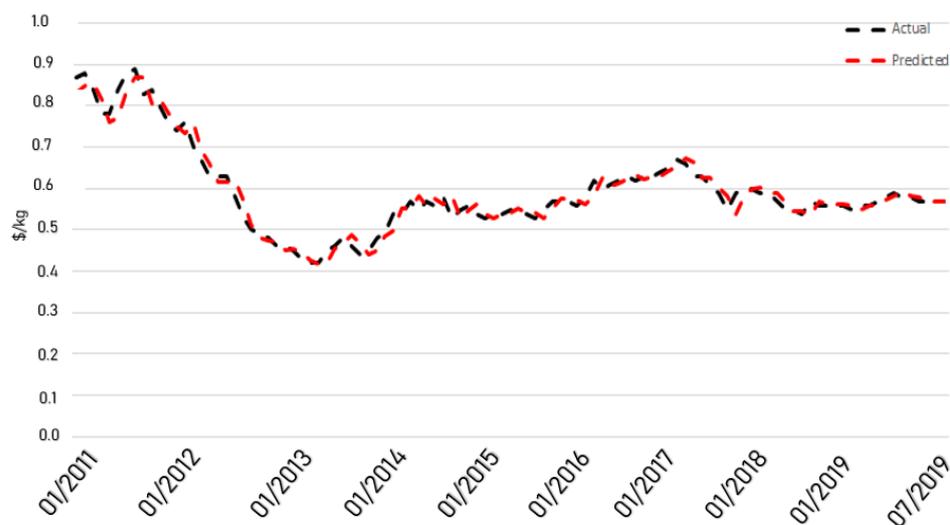


Figure 15. USA actual \times predicted.

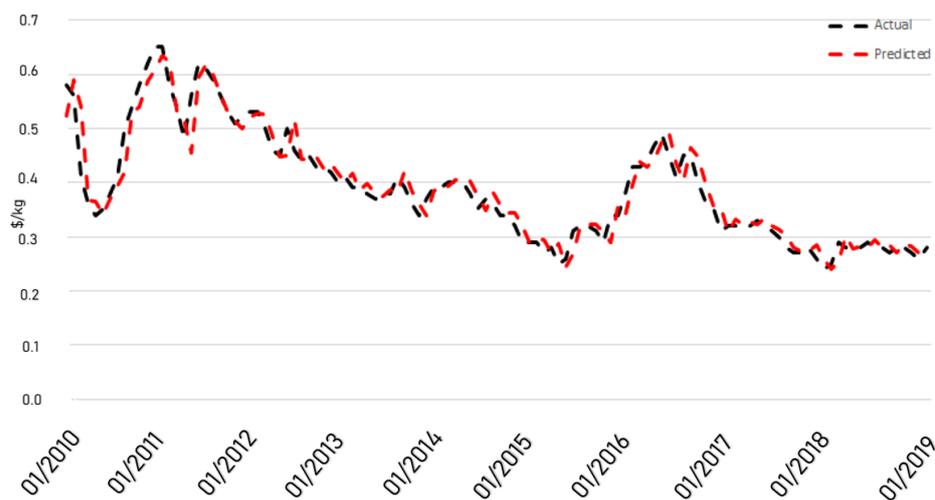


Figure 16. World actual \times predicted.

5. Conclusions

This study investigated the application of several models to predict the sugar prices from four databases: Brazil, the European Union, the United States, and the world databases. This is an essential task, as sugar is present in several sectors of the economy. Also, it presents a fragmented global market, which is evident from its high volatility.

For this purpose, several prediction methods were evaluated: linear methods (AR and ARMA) and Artificial Neural Networks (MLP, ELM, Elman, Jordan, ESN-Jaeger, and ESN-Ozturk et al.). In this case, the Wrapper method selected the number of inputs of each model. Furthermore, different forecast horizons were considered: $P = 1, 3, 6,$ and 12 steps ahead.

In general, the computational results show superior performance for the neural models. This is unsurprising, since real-world problems are often complex, with linear and nonlinear components. The ability of these methods to process nonlinear information is one of the most important advantages of ANN compared to traditional statistical approaches.

For future investigations, it is suggested to investigate the application of combination models, such as Ensembles and Hybrid Models, based on Error Correction. In addition, using deep neural models is an interesting possibility, mainly applying Long Short-Term Memory (LSTM) networks due to the encouraging results in forecasting tasks reported in the literature [57–60]. Finally, a comparison of the findings of this work and the use of equilibrium models should be evaluated.

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

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

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

References

1. OECD-FAO. OECD-FAO Agricultural Outlook 2019. 2023. Available online: <https://www.oecd.org/agriculture/oecd-fao-agricultural-outlook-2019/> (accessed on 1 March 2021).
2. Salehi, R.; Asaadi, M.A.; Rahimi, M.H.; Mehrabi, A. The information technology barriers in supply chain of sugarcane in Khuzestan province, Iran: A combined ANP-DEMATEL approach. *Inf. Process. Agric.* **2021**, *8*, 458–468. [CrossRef]
3. Silva, N.; Siqueira, I.; Okida, S.; Stevan, S.L.; Siqueira, H. Neural Networks for Predicting Prices of Sugarcane Derivatives. *Sugar Tech* **2019**, *21*, 514–523. [CrossRef]
4. Deina, C.; Prates, M.H.d.A.; Alves, C.H.R.; Martins, M.S.R.; Trojan, F.; Stevan, S.L., Jr.; Siqueira, H.V. A methodology for coffee price forecasting based on extreme learning machines. *Inf. Process. Agric.* **2022**, *9*, 556–565. [CrossRef]
5. Maitah, M.; Smutka, L. The Development of World Sugar Prices. *Sugar Tech* **2019**, *21*, 1–8. [CrossRef]
6. der Mensbrugghe, D.V.; Beghin, J.C.; Mitchell, D. *Modeling Tariff Rate Quotas in a Global Context: The Case of Sugar Markets in OECD Countries*; Food and Agricultural Policy Research Institute (FAPRI): Ames, IA, USA, 2003.
7. OECD and Food and Agriculture Organization of the United Nations. *OECD-FAO Agricultural Outlook 2016–2025*; OECD: Paris, France, 2016; 136p. [CrossRef]
8. Rumánková, L.; Smutka, L.; Maitah, M.; Benešová, I. The Interrelationship between Sugar Prices at the Main World Sugar Commodities Markets. *Sugar Tech* **2019**, *21*, 853–861. [CrossRef]
9. Van der Eng, P. *Agricultural Growth in Indonesia: Productivity Change and Policy Impact since 1880*; Springer: London, UK, 1996.
10. Yara Brasil. Produção Mundial de Cana-de-Açúcar. Available online: <https://www.yarabrasil.com.br/conteudo-agronomico/blog/producao-mundial-de-cana-de-acucar/> (accessed on 1 March 2021).
11. OECD and Food and Agriculture Organization of the United Nations. *OECD-FAO Agricultural Outlook 2019–2028*; OECD: Paris, France, 2019; 326p. Available online: https://www.oecd-ilibrary.org/content/publication/agr_outlook-2019-en (accessed on 1 March 2021).
12. USDA, FAS. Sugar: World Markets and Trade. 2020. Available online: <https://fas.usda.gov/data/sugar-world-markets-and-trade> (accessed on 1 March 2021).
13. Lin, Bi.; Variyam, J.N.; Allshouse, J.E.; Cromartie, J. *Food and Agricultural Commodity Consumption in the United States: Looking Ahead To 2020*; Agricultural Economic Reports, Number 33959; United States Department of Agriculture, Economic Research Service: Washington, DC, USA, 2003. Available online: <https://ideas.repec.org/p/ags/uerser/33959.html> (accessed on 1 March 2021).
14. Pagani, V.; Stella, T.; Guarneri, T.; Finotto, G.; Berg, M.v.; Marin, F.R.; Acutis, M.; Confalonieri, R. Forecasting sugarcane yields using agro-climatic indicators and Canegro model: A case study in the main production region in Brazil. *Agric. Syst.* **2017**, *154*, 45–52. [CrossRef]
15. Amrouk, E.M.; Heckelei, T. Forecasting International Sugar Prices: A Bayesian Model Average Analysis. *Sugar Tech* **2020**, *22*, 552–562. [CrossRef]
16. Adenauer, M.; Louhichi, K.; de Frahan, B.H.; Witzke, H.P. Impact of the “Everything but Arms” Initiative on the EU Sugar Sub-Sector. *EcoMod2004*, **2010**, 330600001. Available online: <https://ideas.repec.org/p/ekd/003306/330600001.html> (accessed on 1 March 2021).
17. Nolte, S.; Buysse, J.; Huylenbroeck, G.V. Modelling the Effects of an Abolition of the EU Sugar Quota on Internal Prices, Production, and Imports. In Proceedings of the 114th Seminar, Berlin, Germany, 15–16 April 2010; European Association of Agricultural Economists: Wageningen The Netherlands, 2010. Available online: <https://ideas.repec.org/p/ags/eea114/61346.html> (accessed on 1 March 2021).
18. Stephen Haley. *World Raw Sugar Prices: The Influence of Brazilian Costs of Production and World Surplus/Deficit Measures*; United States Department of Agriculture: Washington, DC, USA, 2013.
19. Haley, S. *Projecting World Raw Sugar Prices*; EMS/USDA: Washington, DC, USA, 2015.

20. Chang, C.; McAleer, M.; Wang, Y. Modelling Volatility Spillovers for Bio-ethanol, Sugarcane and Corn Spot and Futures Prices. *Renew. Sustain. Energy Rev.* **2018**, *81*, 1002–1018. [CrossRef]
21. Rabobank. Sugar Quarterly Q4 2020. Available online: <https://research.rabobank.com/far/en/sectors/sugar/sugar-quarterly-q4-2020.html> (accessed on 1 March 2021).
22. Bisgaard, S.; Kulahci, M. *Time Series Analysis and Forecasting by Example*; John Wiley & Sons: Hoboken, NJ, USA, 2011.
23. Ribeiro, C.; Oliveira, S. A Hybrid Commodity Price-Forecasting Model Applied to the Sugar–Alcohol Sector. *Aust. J. Agric. Resour. Econ.* **2011**, *55*, 180–198. [CrossRef]
24. de Melo, B.; Milioni, A.Z.; Júnior, C.L.N. Daily and Monthly Sugar Price Forecasting Using the Mixture of Local Expert Models. *Pesqui. Oper.* **2007**, *27*, 235–246. [CrossRef]
25. Esam, A.B. Economic Modelling and Forecasting of Sugar Production and Consumption in Egypt. *Int. J. Agric. Econ.* **2017**, *2*, 96–109. [CrossRef]
26. Mehmood, Q.; Sial, M.H.; Riaz, M.; Shaheen, N. Forecasting the Production of Sugarcane in Pakistan for the Year 2018–2030. Using Box-Jenkin’s Methodology. *JAPS, J. Anim. Plant Sci.* **2019**, *29*, 1396–1401.
27. Vishawajith, K.P.; Sahu, P.K.; Dhekale, B.S.; Mishra, P. Modelling and Forecasting Sugarcane and Sugar Production in India. *Indian J. Econ. Dev.* **2016**, *12*, 71–80. [CrossRef]
28. Box, G.E.P.; Jenkins, G.M.; Reinsel, G.C.; Ljung, G.M. *Time Series Analysis: Forecasting and Control*; John Wiley & Sons: Hoboken, NJ, USA, 2015.
29. Santos, D.S.d., Jr.; Neto, P.S.G.d.; Oliveira, J.a.F.L.d.; Siqueira, H.V.; Barchi, T.M.; Lima, A.R.; Madeiro, F.; Dantas, D.A.P.; Converti, A.; Pereira, A.C.; et al. Solar Irradiance Forecasting Using Dynamic Ensemble Selection. *Appl. Sci.* **2022**, *12*, 3510. [CrossRef]
30. Nayak, S.K.; Nayak, S.C.; Das, S. Modeling and Forecasting Cryptocurrency Closing Prices with Rao Algorithm-Based Artificial Neural Networks: A Machine Learning Approach. *FinTech* **2021**, *1*, 47–62. [CrossRef]
31. Neto, P.S.G.d.; Oliveira, J.a.F.L.d.; Júnior, D.S.d.O.S.; Siqueira, H.V.; Marinho, M.H.D.N.; Madeiro, F. A Hybrid Nonlinear Combination System for Monthly Wind Speed Forecasting. *IEEE Access* **2020**, *8*, 191365–191377. [CrossRef]
32. Suresh, K.K.; Priya, S.R.K. Forecasting Sugarcane Yield of Tamilnadu Using ARIMA Models. *Sugar Tech* **2011**, *13*, 23–26. [CrossRef]
33. Kotu, V.; Deshpande, B. (Eds.) Chapter 12–Time Series Forecasting. In *Data Science*, 2nd ed.; Morgan Kaufmann: Burlington, MA, USA, 2019; pp. 395–445. ISBN 978-0-12-814761-0. [CrossRef]
34. Gyamerah, S.A.; Arthur, J.; Akuamoah, S.W.; Sithole, Y. Measurement and Impact of Longevity Risk in Portfolios of Pension Annuity: The Case in Sub Saharan Africa. *FinTech* **2023**, *2*, 48–67. [CrossRef]
35. Siqueira, H.; Luna, I.; Alves, T.A.; Tadano, Y.d. The Direct Connection Between Box & Jenkins Methodology and Adaptive Filtering Theory. *Math. Eng. Sci. Aerosp. (MESA)* **2019**, *10*, 27.
36. 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.* **2006**, *20*, 143–156. [CrossRef]
37. Hong, Y. Hypothesis Testing in Time Series via the Empirical Characteristic Function: A Generalized Spectral Density Approach. *J. Am. Stat. Assoc.* **1999**, *94*, 1201–1220. <http://www.jstor.org/stable/2669935>. [CrossRef]
38. Kosmas, I.; Papadopoulos, T.; Dede, G.; Michalakelis, C. The Use of Artificial Neural Networks in the Public Sector. *FinTech* **2023**, *2*, 138–152. [CrossRef]
39. Belotti, J.; Siqueira, H.; Araujo, L.; Stevan, S.L., Jr.; de Mattos Neto, P.S.; Marinho, M.H.N.; Oliveira, J.F.L.d.; Usberti, F.; Filho, M.d.A.L.; Converti, A.; et al. Neural-based ensembles and unorganized machines to predict streamflow series from hydroelectric plants. *Energies* **2020**, *13*, 4769. [CrossRef]
40. 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]
41. Siqueira, H.; Boccato, L.; Attux, R.; Lyra, C. Unorganized machines for seasonal streamflow series forecasting. *Int. J. Neural Syst.* **2014**, *24*, 1430009. [CrossRef]
42. Haykin, S. *Neural Networks: A Comprehensive Foundation*; Prentice-Hall, Inc.: Hoboken, NJ, USA, 2007.
43. Siqueira, H.; Luna, I. Performance comparison of feedforward neural networks applied to streamflow series forecasting. *Math. Eng. Sci. Aerosp. (MESA)* **2019**, *10*, 41.
44. Huang, G.; Zhu, Q.; Siew, C. Extreme learning machine: A new learning scheme of feedforward neural networks. In Proceedings of the 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No.04CH37541), Budapest, Hungary, 25–29 July 2004; Volume 2, pp. 985–990.
45. Yu, H.; Yang, L.; Li, D.; Chen, Y. A hybrid intelligent soft computing method for ammonia nitrogen prediction in aquaculture. *Inf. Process. Agric.* **2021**, *8*, 64–74. [CrossRef]
46. Tadano, Y.d.; Siqueira, H.V.; Alves, T.A. Unorganized machines to predict hospital admissions for respiratory diseases. In Proceedings of the 2016 IEEE Latin American Conference on Computational Intelligence (LA-CCI), Cartagena, Colombia, 2–4 November 2016; pp. 1–6.
47. Huang, G.; Zhu, Q.; Siew, C. Extreme Learning Machine: Theory and Applications. *Neurocomputing* **2006**, *70*, 489–501. [CrossRef]
48. Nahvi, B.; Habibi, J.; Mohammadi, K.; Shamshirband, S.; Razgan, O.S.A. Using self-adaptive evolutionary algorithm to improve the performance of an extreme learning machine for estimating soil temperature. *Comput. Electron. Agric.* **2016**, *124*, 150–160. [CrossRef]

49. Jaeger, H. *The “Echo State” Approach to Analysing and Training Recurrent Neural Networks—with an Erratum Note*; GMD Technical Report; German National Research Center for Information Technology: Bonn, Germany, 2001; Volume 148, p. 13.
50. Ozturk, M.C.; Xu, D.; Principe, J.C. Analysis and Design of Echo State Networks. *Neural Comput.* **2007**, *19*, 111–138. [[CrossRef](#)]
51. Belotti, J.; Mendes, J.J.; Leme, M.; Trojan, F.; Stevan, S.L.; Siqueira, H. Comparative Study of Forecasting Approaches in Monthly Streamflow Series from Brazilian Hydroelectric Plants using Extreme Learning Machines and Box & Jenkins Models. *J. Hydrol. Hydromech.* **2021**, *69*, 180–195.
52. The World Bank. Commodity Markets–Sugar Price. 2019. Available online: <https://www.worldbank.org/en/research/commodity-markets> (accessed on 1 March 2021).
53. CEPEA, Center for Advanced Studies in Applied Economy. Sugar Price-Brazil. 2019. Available online: <http://cepea.esalq.usp.br> (accessed on 21 August 2019).
54. Hyndman, R.J.; Athanasopoulos, G. *Forecasting: Principles and Practice*; OTexts: Melbourne, Australia, 2021
55. Guerreiro, M.T.; Guerreiro, E.M.A.; Barchi, T.M.; Biluca, J.; Alves, T.A.; Tadano, Y.d.; Trojan, F.; Siqueira, H.V. Anomaly detection in automotive industry using clustering methods—A case study. *Appl. Sci.* **2021**, *11*, 9868. [[CrossRef](#)]
56. Jevšenak, J.; Levanič, T. Should artificial neural networks replace linear models in tree ring based climate reconstructions? *Dendrochronologia* **2016**, *40*, 102–109. [[CrossRef](#)]
57. Morid, M.A.; Sheng, O.R.L.; Dunbar, J. Time Series Prediction Using Deep Learning Methods in Healthcare. *Assoc. Comput. Mach.* **2023**, *14*, 1–29. [[CrossRef](#)]
58. Zhang, H.; Wang, C.; Tian, S.; Lu, B.; Zhang, L.; Ning, X.; Bai, X. Deep learning-based 3D point cloud classification: A systematic survey and outlook. *Displays* **2023**, *79*, 102456. [[CrossRef](#)]
59. Sharifani, K.; Amini, M. Machine Learning and Deep Learning: A Review of Methods and Applications. *World Inf. Technol. Eng. J.* **2023**, *10*, 3897–3904.
60. Wang, C.; Ning, X.; Sun, L.; Zhang, L.; Li, W.; Bai, X. Learning discriminative features by covering local geometric space for point cloud analysis. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–15. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.