Forecast modeling a time series of water reservoir levels using exponential smoothing method

Lydiane F. de Souza
Center for Exact Sciences and Technologies
Federal University of Western Bahia
Barreiras, Brazil
lydianede1@hotmail.com

***Abstract*—Exponential smoothing (ES) is a time series forecasting method that presents the forecast based on trend and seasonality components. In this work, we study the behavior of two time series that describe the level of the water reservoirs of the Descoberto and Santa Maria dams. We trained the fifteen models present in the Pregels taxonomy, the criterion for choosing the model consists of the model with the lowest Akaike information criterion. The results indicate that the exponential smoothing model with damped additive trend and additive seasonality best describes both time series.**

***Index Terms*—Exponential smoothing; Forecast model; time series**

I. Introduction

Exponential smoothing (ES) methods are widely used in demand forecasting [1], production and inventory in the business area [2]–[4], telecommunications data [5] and is also used in weather forecast [6], [7]. Concerning the study of time series related to the level of reservoirs, there are some works related to hydrology using forecasting methods [8]–[11], the use of imputation models, variable selection models to improve the forecasting capacity [12], and in more recent works the use of deep learning [13]. The ES forecasting methods have been widely used since 1950 because their mathematical formulation is simple, requires little computational time, and results with reasonable accuracy. The ES methods have 15 variations, depending on the trend and seasonality used in the forecast; the best known are simple exponential smoothing [14], Holt method [15], and Holt-Winters method [16].

This work will use the ES methods to obtain the best prediction model of the historical series of water levels of the two main water reservoirs that supply the Federal District in Brazil. The Descoberto and Santa Maria dams reservoirs are the oldest water reservoirs in the Federal District. The historical series of the level of the Descoberto and Santa Maria reservoirs used in this research starts on April 1987 and ends on October 2021 [17]; the historical series of the Descoberto dam has 12307 data, and the Santa Maria dam has 12399 data.

The study aims to obtain the most suitable model for each reservoir and predict this series's behavior for the coming years to encourage public policies for preserving water resources. We will get the best model based on the lowest value of Akaike's Information Criterion (AIC) [18], [19]. In this work, we will see that for the Descoberto dam and the Santa Maria dam, the model with the lowest AIC has additive-damped trends and additive seasonality.

In section 2 of this work, we will describe the exponential smoothing models and how to calculate the AIC of the models. Section 3 will present this study's historical series of reservoir levels. In section 4, we will calculate and analyze the results of the AIC and BIC criteria of each ES model; we will obtain the appropriate model for each of the historical series, the forecast of these models for the series, and the residual analysis, to confirm if the chosen model is adequate. Finally, section 5 presents the conclusions and perspectives of this work.

II. Exponential Smoothing

The data for a time series are y_1, y_2, ..., y_T. A simple method of forecasting this series is to consider the forecast equal to the last observation:

$$F_{T+h} = y_T. \tag{1}$$

Moreover, we assume that the most recent value is the most important, such that F_{T+h} is the prediction of the variable y for all times above T, with $h = 1, 2, 3, ...$. The average method assumes that all observations are of equal importance and the data have equal weights when generating predictions:

$$F_{T+h} = \frac{1}{T}\sum_{t=1}^{T} y_T. \tag{2}$$

Simple exponential smoothing (SES) is an intermediate forecasting method where the forecast weights decrease exponentially and forecasts for $h > 1$ always have the same value, that is:

$$F_{T+h} = F_{T+1} = \alpha y_T + (1-\alpha)y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + ..., \tag{3}$$

α is the smoothing parameter, $0 \leq \alpha \leq 1$ and the closer to 1 the parameter α is, the greater the weight of recent observations. SES has a greater weight for more recent observations than for past times in time series without trend and seasonality [20]. Another way to represent the forecast is to define it through the level l_T, such that:

$$F_{T+h} = F_{T+1} = l_T \tag{4}$$

$$l_T = \alpha y_T + (1-\alpha)l_{T-1}. \tag{5}$$

In 1957 Charles Holt proposed a method for forecasting the time series with the trend. The forecast is through:

$$F_{T+h} = l_T + hb_T \tag{6}$$
$$l_T = \alpha y_T + (1-\alpha)(l_{T-1} + b_{T-1}) \tag{7}$$
$$b_T = \beta(l_T + b_{T-1}) + (1-\beta)b_{T-1}, \tag{8}$$

with l_T the level at time T, b_T the trend or slope of the series at time T, α is the level smoothing parameter, $0 \leq \alpha \leq 1$, β is the smoothing parameter for the trend and $0 \leq \beta \leq 1$. The Holt-Winters method (1960) is an extension of the Holt method for series with seasonality; its forecast is calculated through l_T, b_T, of the seasonality at the instant T, s_T, and the smoothing parameters are α, β and γ, so:

$$F_{T+h} = l_T + hb_T + s_{T+h-m(k+1)} \tag{9}$$
$$l_T = \alpha(y_T - s_{T-m}) + (1-\alpha)(l_{T-1} + b_{T-1}) \tag{10}$$
$$b_T = \beta(l_T - l_{T-1}) + (1-\beta)b_{T-1} \tag{11}$$
$$s_T = \gamma(F_T - l_{T-1} - b_{T-1}) + (1-\gamma)s_{T-m}, \tag{12}$$

where m is the frequency of seasonality, that is, the number of seasons in the year [21], and k is the integer part of $(h-1)/m$. This is the additive seasonality version used when the seasonality is constant, and there is the multiplicative seasonality version when the seasonality varies with the level.

In addition to the forecasting methods mentioned above, there is a taxonomy proposed by Pegels [22] and employed by Hyndman [23], which identifies and differentiates forecasting methods according to their trend and seasonality. The Table I identifies the fifteen forecasting methods tested in this work.

Trend	Seasonality		
	N-none	A-additive	M-multiplicative
N-none	NN	NA	NM
A-additive	AN	AA	AM
M-multiplicative	AM	MA	MM
AD-additive damped	ADN	ADA	ADM
MD-multiplicative damped	MDN	MDA	MDM

TABLE I
TAXONOMY EXPONENTIAL SMOOTHING METHODS

The method we choose to get the best model for prediction is by calculating Akaike's Information Criterion (AIC) and also called the Schwarz criterion:

$$AIC = T \log\left(\frac{SSE}{T}\right) + 2c, \tag{13}$$

with T the total number of data used to estimate or train the model, $SSE = \sum_{t=1}^{T} e_t^2$ the sum of squared errors, $e_t = y_t - \hat{y}_t$, where $\hat{y}_t$ are the model predictions and c the total number of estimated parameters in the model, $c = m \times n_s + 2 \times n_t + 2 + n_d$. The variables n_s, n_t and n_d can have values of 0 or 1 depending on whether or not the model has seasonality, trend and damping, respectively. For instance, the variable n_s is 0 when the model has no seasonality and $n_s = 1$ when the model has seasonality.

The model chosen will have the smallest AIC since the smaller the AIC, the less information is lost from the model. The AIC depends not only on the SSE but also on the total number of estimated parameters. The choice of AIC instead of mean absolute percentage error (MAPE) or mean squared error (MSE) is because these error measures do not consider the number of parameters predicted by the model—the AIC penalizes models with many parameters to the detriment of models with a lower c [21].

III. Reservoir water level data

The reservoirs of the Descoberto, Santa Maria, and Lake Paranoá dams are the main ones responsible for supplying the Federal District. Together, the first two dams provide 83% of the population [24].

Obtain the data in [17], have the average daily level of the reservoirs of the Descoberto dam, with 12307 data, of the Santa Maria dam, with 12399 data, and of the Paranoá lake, with 1511 data. However, the small amount of data from Lake Paranoá compared to the other two dams restricts the study to the Descoberto and Santa Maria dams, with historical series from April 1987 to October 2021. Therefore, the historical series of the Descoberto reservoir is shown in Fig. 1, while the historical series of the Santa Maria reservoir is in Fig. 2; in addition, in 2017, there was the lowest average level recorded for the two dams.

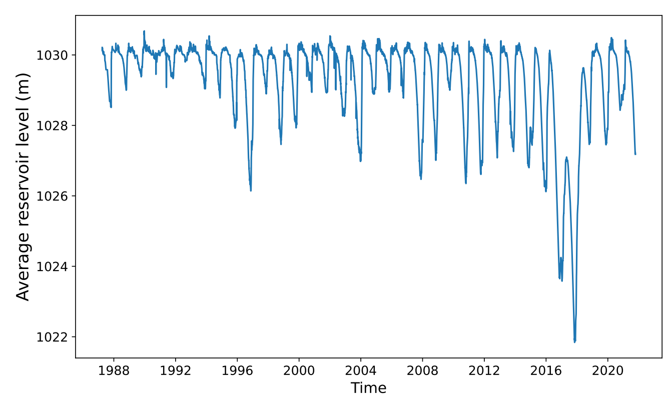

Fig. 1. Historical Series of the Descoberto Dam Reservoir (April, 1987-October, 2021)

The Descoberto reservoir reached the minimum level of 1021.84 m on November 07, 2017, and the Santa Maria reservoir reached the minimum level of 1064.17 m on November 25, 2017. As we can see in the heat map of the two dams Fig. 3 and 4, in 2017, there was a dry period, presenting atypical behavior throughout the year.

IV. Results and discussions

We train the prediction models indicated in Table I for the datasets of the Descoberto Dam and the Santa Maria Dam; for both time series, we trained ninety percent of the data, and ten percent were for testing. The choice of this percentage is because the series presents this period of extreme drought between 2016 and 2018, such that in this period, the series level suffers an abrupt decline, interfering with the test points.

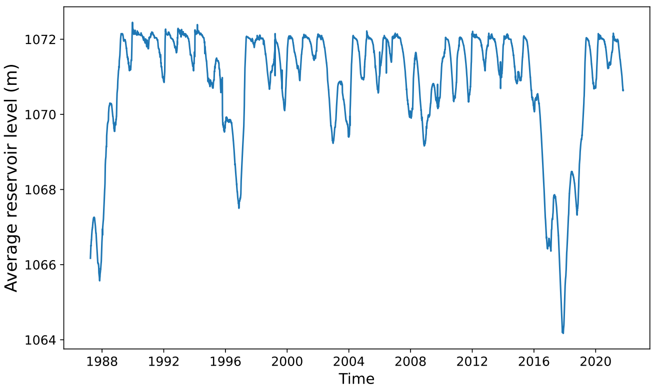

Fig. 2. Historical Series of the Santa Maria Dam Reservoir (April, 1987-October, 2021)

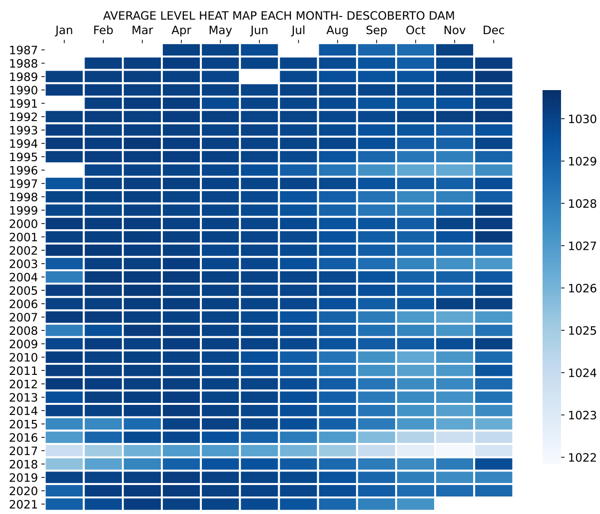

Fig. 3. Average level heat map each month- Descoberto Dam (April, 1987-October, 2021)

Once the AIC is calculated, for each of the models, we obtain Table II for the time series of the Descoberto Dam and Table III for the Santa Maria Dam.

The time series studied in this work show seasonality; for this reason, we will choose models with seasonality. Therefore, the lowest AIC in ascending order are ADA, AA, MA, and NA for the Descoberto Dam. As for the Santa Maria Dam, the models with the lowest AIC in ascending order are ADA, AA, NA, and AM. After choosing the four models with the lowest AIC for each time series, we can see in Figs. 3 and 4 that 2017 was an atypical period and presented drought. The presence of these data in the model's training can interfere with the choice of the appropriate model, for that we performed a second training of the data before the dry period. In this step,

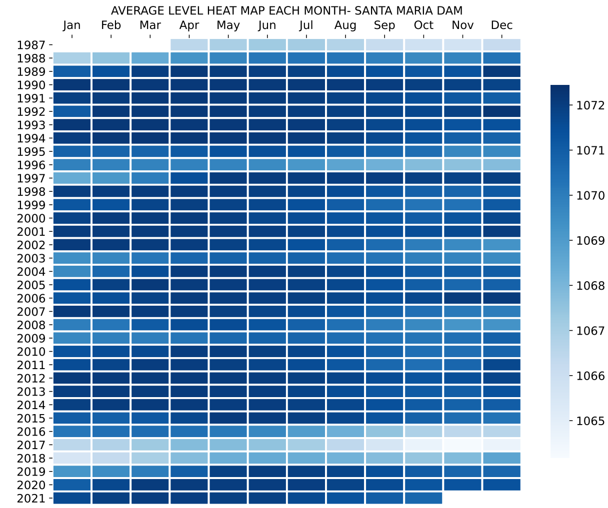

Fig. 4. Average level heat map each month- Santa Maria Dam (April, 1987-October, 2021)

we calculated the AIC in only four exponential smoothing models that performed better in the first test, and we trained eighty-four percent of the data. We obtained the results from Table IV for the Descoberto Dam and Table V for the Santa Maria Dam. We can conclude that the model with the lowest AIC is the ADA model for both series.

The ADA model is an exponential smoothing model with additive damped trend and additive seasonality, whose recursive formula is:

$$F_{T+h} = l_T + \phi_h b_T + s_{T+h-m(k+1)} \tag{14}$$

$$l_T = \alpha(y_T - s_{T-m}) + (1-\alpha)(l_{T-1} + \phi b_{T-1}) \tag{15}$$

$$b_T = \beta(l_T - l_{T-1}) + (1-\beta)\phi b_{T-1} \tag{16}$$

$$s_T = \gamma(F_T - l_{T-1} - \phi b_{T-1}) + (1-\gamma)s_{T-m} \tag{17}$$

$$\phi_h = \phi + \phi^2 + ... + \phi^h, \tag{18}$$

with ϕ the damping parameter, $0 < \phi < 1$.

Considering the data from the Descoberto Dam and training ninety percent of the data, we obtain that the AIC value of the ADA model is -67469.059 and the smoothing parameters are:

$$\alpha = 0.931 \tag{19}$$

$$\beta = 0.098 \tag{20}$$

$$\gamma = 0.002 \tag{21}$$

$$\phi = 0.978, \tag{22}$$

with $m = 365$, $h \in [0, 1172]$ e $T = 11136$. Figure 5 shows the historical series of the Descoberto Dam and the ADA forecast model for the ten percent of the data selected and tested. For the data from the Santa Maria Dam and training ninety percent

Trend	Seasonality		
	N-None	A-Additive	M-Multiplicative
N-None	-66702.198	-66915.802	-65487.895
A-Additive	-67801.853	-67370.276	-65961.216
M-Multiplicative	-67801.657	-67369.428	-65938.873
AD-Additive Damped	-67933.668	-67469.059	-66099.082
MD-Multiplicative Damped	-67933.873	-7815.440	-66067.551

TABLE II
AIC VALUES FOR THE EXPONENTIAL SMOOTHING MODELS APPLIED TO THE TIME SERIES OF THE DESCOBERTO DAM

Trend	Seasonality		
	N-None	A-Additive	M-Multiplicative
N-None	-76368.183	-75995.777	-72954.198
A-Additive	-77546.878	-76029.366	-73187.861
M-Multiplicative	77726.449	-55199.861	-65738.916
AD-Additive Damped	-77814.129	-77317.712	-73054.512
MD-Multiplicative Damped	-77814.209	136021.169	-73097.398

TABLE III
AIC VALUES FOR THE EXPONENTIAL SMOOTHING MODELS APPLIED TO THE TIME SERIES OF THE SANTA MARIA DAM

Models	AIC
ADA	-62334.936
AA	-62232.353
MA	-62231.139
NA	-61962.861

TABLE IV
AIC VALUES FOR THE ADA, AA, MA AND NA MODELS OF THE DESCOBERTO DAM, TRAINING EIGHTY-FOUR PERCENT OF THE DATA

Models	AIC
ADA	-71364.011
AA	-70546.029
NA	-70482.079
AM	-68057.242

TABLE V
AIC VALUES FOR THE ADA, AA, NA AND AM MODELS OF THE SANTA MARIA DAM, TRAINING EIGHTY-FOUR PERCENT OF THE DATA

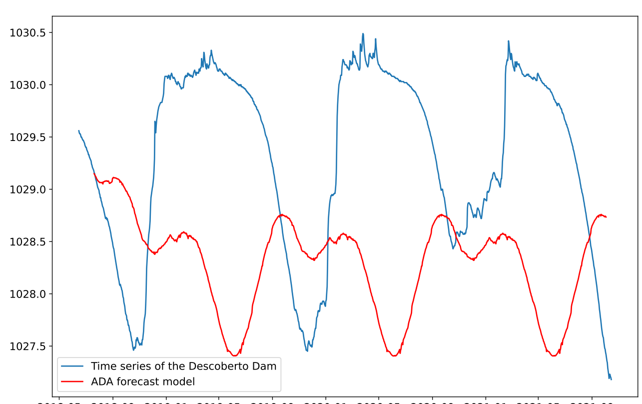

Fig. 5. Historical Series of the Descoberto Dam Reservoir (May, 2018 - October, 2021) and ADA forecast model.

of the data, we obtain that the AIC value of the ADA model is -77317.712 and the smoothing parameters are:

$$\alpha = 0.774 \quad (23)$$

$$\beta = 0.060 \quad (24)$$

$$\gamma = 0.002 \quad (25)$$

$$\phi = 0.991, \quad (26)$$

with $m = 365$, $h \in [0, 1240]$ e $T = 11159$. Figure 6 shows the historical series of the Santa Maria Dam and the ADA forecast model for the ten percent of the data selected and tested.

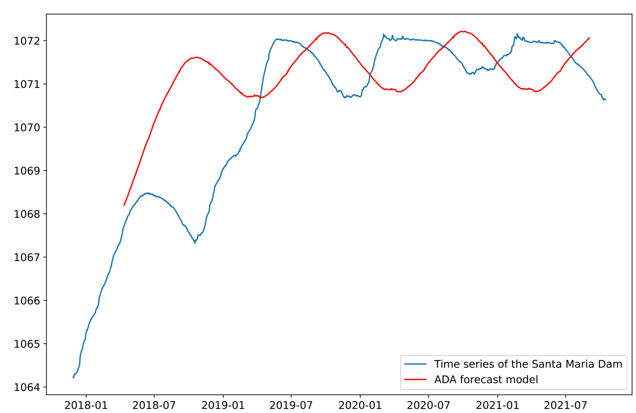

Fig. 6. Historical Series of the Santa Maria Dam Reservoir (January, 2018 - October, 2021) and ADA forecast model.

In this work, we use the residual histogram method and the Normal Q-Q plot [25] to verify that the set of residuals of the model is normality. We draw in Figure 7 the residue histograms for the Descoberto and Santa Maria dams. The histograms show that the model used for both time series is normalized and corresponds to a Gaussian distribution. Furthermore, in the Normal Q-Q plot method graphs for the two dams in Figure 8, we observed the residuals in the graph fall approximately along a straight diagonal line, i.e., the residuals are normality. Therefore, we conclude that the ADA model for the time series of the Descoberto and Santa Maria dams is normal.

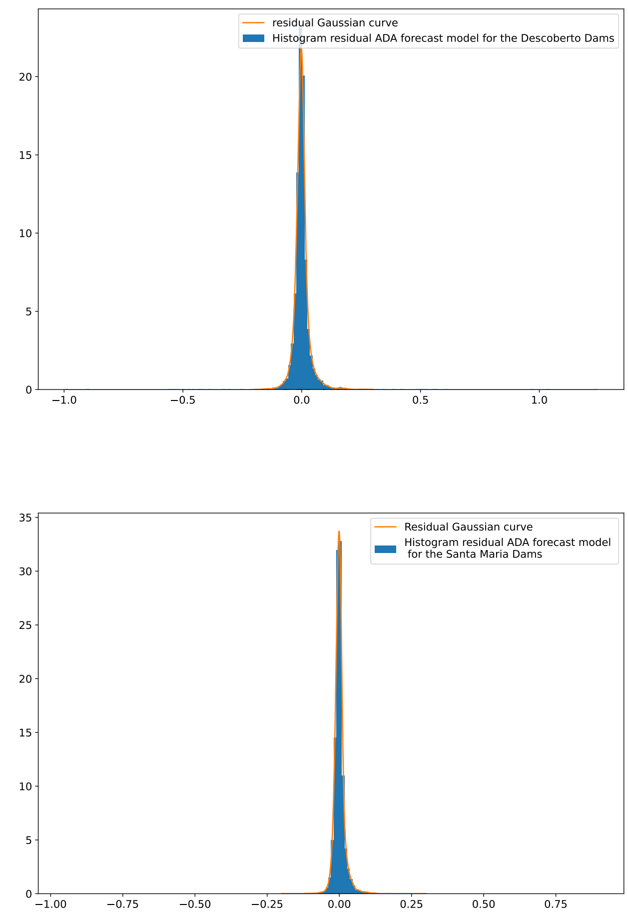

Fig. 7. *Above:* Histogram of residues for the Descoberto Dams. *Below:* Histogram of residues for the Santa Maria Dams.

V. Conclusion

In this work, we study the exponential smoothing models proposed by the Pregels taxonomy, which can present forecasts based on the seasonal component, trend, and level.

We used the historical series of water levels from the Descoberto and Santa Maria reservoirs, trained for each exponential smoothing model, and calculated the AIC value to obtain the models with the lowest AIC. We performed a new test training the data before the 2017 drought to verify whether the model chosen for the two dams suffered any changes due to the dry period. We confirmed that the model with the lowest AIC in both tests is the ADA model.

Therefore, this study indicates that the ADA model, an exponential smoothing model with damped additive trend and

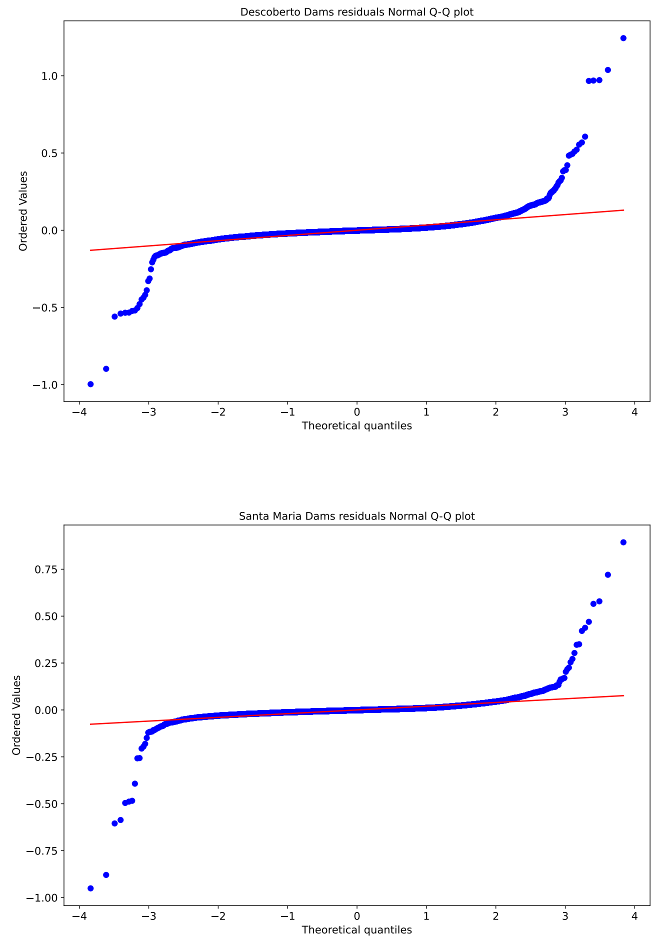

Fig. 8. *Above:* Normal Q-Q plot of residues for the Descoberto Dams. *Below:* Normal Q-Q plot of residues for the Santa Maria Dams.

additive seasonality, is the appropriate model to describe the behavior of the two-time series. Based on this model, we also analyzed the residuals to verify whether the model is normalized and confirmed this hypothesis. The idea is to work with more robust forecasting techniques in future works, as well as to compare the performance of the models in order to obtain a minor error between the actual and predicted values [13], as is the case of recurrent neural networks (RNNs), seasonal autoregressive integrated moving average (SARIMA), specifically memory (LSTM), which is used not only in forecasting time series but also for classifying and processing them. [26].

References

[1] Gardner Jr., E.S., (1985). Exponential smoothing: The state of the art. Journal of Forecasting, 4, 1-28.

[2] Siregar, B., Butar-Butar, I. A., Rahmat, RF., Andayani, U. & Fahmi, F. (2017). Comparison of Exponential Smoothing Methods in Forecasting Palm Oil Real Production. Journal of Physics: Conference Series, 801, 012004.

[3] Tratara, L. F. & Strmčnikb, E. (2016). The comparison of Holt-Winters method and Multiple regression method: A case study. Energy, 109, 266-276.

[4] Pamungkas, A., Puspasari, R. and Nurfiarini, A., Zulkarnain, R. Waryanto, W. (2021). Comparison of Exponential Smoothing Methods for Forecasting Marine Fish Production in Pekalongan Waters, Central Java. IOP Conference Series: Earth and Environmental Science, 934, 012016.

[5] Gardner Jr., E. S. & Diaz-Saiz, J. (2008). Exponential smoothing in the telecommunications data. International Journal of Forecasting, 24, 170-174.

[6] Mahajan, S.,Chen, L-J. & Tsai,T.-C. (2018).Short-Term PM2.5 Forecasting Using Exponential Smoothing Method: A Comparative Analysis. Sensors, 18(10), 3223.

[7] Zafar, S., Kashif, M., Khan, M. & Nida, H. (2021).Application of Simple Exponential Smoothing Method for Temperature Forecasting in Two Major Cities Of The Punjab, Pakistan. Agrobiological Records, 4(01),64-68.

[8] F.-J. Chang and Y.-T. Chang, "Adaptive neuro-fuzzy inference system for prediction of water level in reservoir," Advances in Water Resources, vol. 29, no. 1, pp. 1–10, 2006.

[9] M. K. Tiwari and C. Chatterjee, "Development of an accurate and reliable hourly flood forecasting model using wavelet-bootstrap-ANN (WBANN) hybrid approach," Journal of Hydrology, vol. 394, no. 3-4, pp. 458–470, 2010.

[10] C.-S. Chen, Y.-D. Jhong, T.-Y. Wu, and S.-T. Chen, "Typhoon event-based evolutionary fuzzy inference model for flood stage forecasting," Journal of Hydrology, vol. 490, pp. 134–143, 2013.

[11] F.-J. Chang, P.-A. Chen, Y.-R. Lu, E. Huang, and K.-Y. Chang, "Real-time multi-step-ahead water level forecasting by recurrent neural networks for urban flood control," Journal of Hydrology, vol. 517, pp. 836–846, 2014.

[12] Jun-He Yang, Ching-Hsue Cheng, Chia-Pan Chan, "A Time-Series Water Level Forecasting Model Based on Imputation and Variable Selection Method", Computational Intelligence and Neuroscience, vol. 2017, Article ID 8734214, 11 pages, 2017.

[13] Atashi, V.; Gorji, H.T.; Shahabi, S.M.; Kardan, R.; Lim, Y.H. Water Level Forecasting Using Deep Learning Time-Series Analysis: A Case Study of Red River of the North. Water 2022, 14, 1971.

[14] Brown, R. G. (1959). Statistical forecasting for inventory control. New York: McGraw Hill.

[15] Holt, C. C. (2004). Forecasting Trends and Seasonal by Exponentially Weighted Averages. International Journal of Forecasting, 20, 5–13.

[16] Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. Management Science, 6, 324 – 34.

[17] Agência Reguladora de Águas, Energia e Saneamento Básico do Distrito Federal, Monitoramentos dos Níveis de reservatórios, https://www.adasa.df.gov.br/monitoramento/niveis-dos-reservatorios, Accessed October 15, 2021.

[18] Kolassa, S. (2011). Combining exponential smoothing forecasts using Akaike weights. International Journal of Forecasting, 27, 238–251.

[19] Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19, 716–723.

[20] Ostertagova, Eva & Ostertag, Oskar. (2012). Forecasting Using Simple Exponential Smoothing Method. Acta Electrotechnica et Informatica, 12, 62–66, 10.2478/v10198-012-0034-2.

[21] Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: Principles and Practice. (2nd ed.) OTexts.

[22] Pegels, C. C. (1969). Exponential forecasting: some new variations. Management Science, 12, 311–315.

[23] Hyndman, R. J., Koehler, A. B., Snyder, R. D. & Grose, S.(2002). A state space framework for automatic forecasting using exponential smoothing methods. International Journal of Forecasting, 18, 439–454.

[24] Agência Nacional de Águas, https://www.ana.gov.br/sar/outros-sistemas-hidricos/df, Accessed December 19, 2022.

[25] Wilk, M. B., & Gnanadesikan, R. (1968). Probability Plotting Methods for the Analysis of Data. Biometrika, 55(1), 1–17.

[26] Song, X., Liu Y., Xue, L., Wang, J., Zhang,J.,Wang, J., Jiang, L., Cheng,Z. (2020). Time-series well performance prediction based on Long Short-Term Memory (LSTM) neural network model. Journal of Petroleum Science and Engineering, 186, 106682.