Utilization of Whittaker-Henderson Smoothing Method for Improving Neural Network Forecasting Accuracy

Hans Pratyaksa[1], Adhistya Erna Permanasari[2*], Silmi Fauziati[3]

Abstract— Health institutions need to ensure the availability of drug stocks for patients. There are challenges related to the uncertainty of the amount of drug use for the next period. Uncertainty can be reduced by analysing historical drug data to predict future demand. Time series can contain spikes or fluctuation pattern which spikes can disguise the main information. Hence, it can affect the accuracy of the prediction model. One widely used forecasting method in the time series data is the artificial neural network (ANN) method. The ANN method requires the pre-processing stage of the data before the training process. The pre-processing stage is essential to obtain information or knowledge. This study focused on applying smoothing methods at the pre-processing stage of the ANN method. The application of the smoothing method was expected to improve the quality of ANN learning data that would lead to better predictive accuracy. This research focuses on implementing the smoothing method in data pre-processing step for ANN method. Smoothing methods used in this research were exponential smoothing (ES) and Whittaker-Henderson (WH) smoothing applied to two time series datasets. The refining method used in this study was the WH method, which was tested on two time series datasets of medicine. The results show that the mean square error (MSE) obtained by applying the WH method was lower than the non-smoothing ANN for both datasets. Evaluation results revealed that implementing WH smoothing method in data pre-processing step for ANN (WH+ANN) provided MSE significantly lower than ANN results with a confidence level of 94% for dataset 1 and 85% for the dataset 2.

Keywords—Forecasting, Pre-processing, Smoothing Method, Whittaker-Henderson, Artificial Neural Network.

I. INTRODUCTION

Medication is an essential requirement for patients as it is one of the primary means used by health experts to treat diseases. Consequently, the hospital or pharmacy must always ensure the availability of medication prescribed to patients in need. Drug availability is one of the challenges for health institutions to provide quality care; therefore, health institutions need proper control of drug availability to avoid adverse effects [1], [2].

An ability to predict future needs is essential in anticipating the need for resource availability. Accurate forecasting enables the decision-makers to anticipate service needs and make decisions regarding the purchase management of resources and supplies over time [1], [3]-[6]. Forecasting can be carried out by intuition or by analyzing time series data. For forecasting based on time series data analysis, the estimator uses time series data patterns that represent the number of requests for time series.

In data mining, many studies related to forecast have been conducted. The Improvement in various method or a combination of several methods have been presented to generate accurate predictions. One method for predicting time series data is the neural network (NN) method. Studies related to the improvement of artificial neural network (ANN) predictive accuracy present various methods, including a combination of optimization methods, such as particle swarm optimization (PSO) or genetic algorithm (GA), with the ANN method [7], [8]; or a combination of two forecasting methods, for example the combination of the autoregressive integrated moving average (ARIMA) method and the ANN [9].

Apart from the previously mentioned approaches, recent studies show that the improvement of ANN methods predictive accuracy can be made by focusing on the pre-processing stage. In the time series data, fluctuating patterns often occur, which can disguise the information value or the main phenomena so that it can degrade the data quality. Therefore, the pre-processing stage is essential to extract important information that influences the performance of forecasting accuracy [10]. One of the pre-processing stages for time series data is data smoothing. The use of smoothing techniques in time series data is expected to extract the primary trend or pattern in the data [11]. Smoothing can also be used to eliminate noise signals [12] or fluctuating data [11], but it does not eliminate essential information on the data [13].

As the ANN predictive accuracy improvement techniques develop, the smoothing methods also improve. Reference [14] has mentioned that one of the reliable smoothing methods in time series data smoothing is the Whittaker-Henderson (WH) smoothing method since it considers the fidelity and smoothness of data smoothing [14], [15]. In addition, [14] has explained and developed an algorithm from the WH method based on generalized cross-validation (GCV). In the study, the proposed WH smoothing method [14] was applied to the pre-processing stage of the ANN method to predict the amount of medicine used by patients.

It is crucial to conduct research related to the use of the WH smoothing method to determine whether it can improve data quality to better the forecasting accuracy of the ANN method. In this study, two data smoothing methods to predict the amount of medicine usage namely the data transformation method WH smoothing method [14] and the exponential smoothing (ES) method [16] were utilized. Both forecasting accuracy results were compared and evaluated.

[1,2,3] *Department of Electrical and Information Engineering, Faculty of Engineering, Universitas Gadjah Mada, Jln. Grafika No. 2, Kampus UGM, Yogyakarta, 55281, INDONESIA (phone: 0274-552305; email: [1]hans_ti10@mail.ugm.ac.id, [2]adhistya@ugm.ac.id, [3]silmi@ugm.ac.id)*
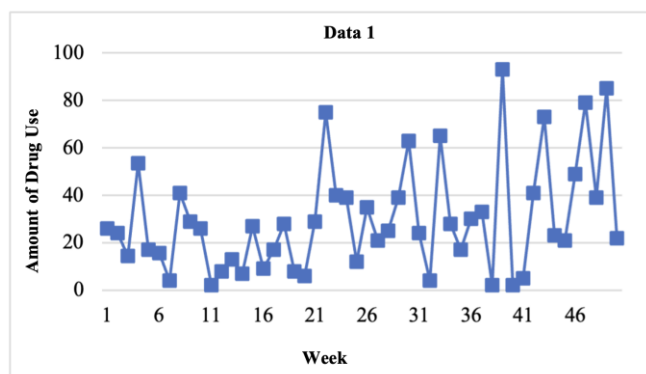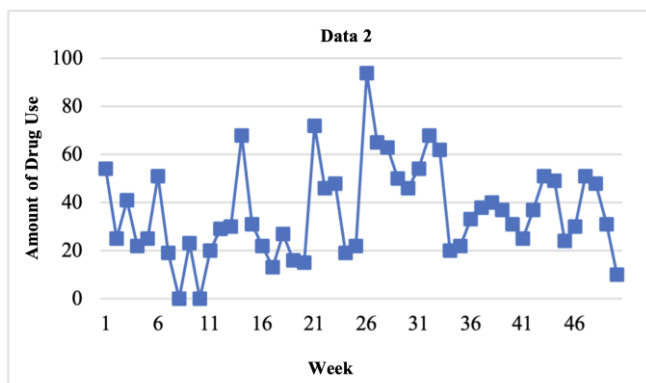*Corresponding Author*

Fig. 1 Time series 1.

Fig. 2 Time series 2.

The rest of the paper is structured as follows. Section 2 introduces related works. Section 3 describes data used in this paper. Section 4 shows methods. Section 5 presents experimental scenarios, while Section 6 reports experimental results. Finally, Section 7 concludes the overall results of the study.

II. Neural Network Forecasting

Forecasting is an activity to predict a future event based on knowledge collected from events that have occurred. Forecasting can be done on time series data. Time series data is a collection of data measured successively at different time intervals [17]. Time series data can be used for various purposes, one of which is to predict future events and to understand an upcoming phenomenon [18]. Prediction of time series data has been used and studied in various fields. There are already various methods for predicting future events in time series data, from linear and nonlinear models.

The development of studies related to time series forecasting suggests that non-linear forecasting models can be conducted. One non-linear model for time series forecasting is using neural network methods [19], [20]. Machine learning is considered to produce better predictive results. There are many approaches to improve the forecasting accuracy of neural network methods.

Neural network modeling is modeling with a black box system that is entirely dependent on input and output data. For this reason, the quality and distribution of sample sets are essential for generalizing the network [16]. Accordingly, the pre-processing stage plays a critical role in neural network

modeling. The pre-processing stage is a crucial step to extract the important information so that it will affect the performance of forecasting accuracy [10]. One of the pre-processing stages for time series data is data smoothing. The use of smoothing techniques in time series data is expected to extract the main trends or patterns in the data [11]. The smoothing can also be used to eliminate unwanted noise signals [12] or inappropriate data [11], but it does not eliminate important information on the data [13]. Therefore, the pre-processing stage by applying the smoothing method can improve the quality of time series data.

A study related to the effect of pre-processing on predictive accuracy has been conducted [21]. In this study, the effect of time series data smoothing in the pre-processing stage on multilayer perception neural network (MLP NN) was studied. The smoothing methods compared were seasonal adjustment series (SAS), seasonal trend cycle (STC), ES (simple, linear, Holt, and Winter), and Box-Jenkins (moving average, autoregression, and ARIMA) methods. The results show that smoothing in the pre-processing stage can improve the forecasting accuracy of NN and the use of smoothing methods is useful to eliminate the noise that occurs.

Reference [21] has proven that one of the time series pre-processing data stages, namely the data smoothing stage applied before the ANN input, could improve the predictive accuracy. Moreover, a related study on improving the accuracy of data forecasting of time series has been carried out [22]. The study focused on data pre-processing to extract redundant information from signals and transforms the data in the pre-processing stage using discrete cosine transform (DCT) to improve the forecasting accuracy of the feedforward neural network (FFNN).

In [16], the ES method was applied to transform time series data used as inputs of the neural network. The results indicate an increase in the forecasting accuracy of gold prices. The study shows that the ES method can be used to improve the data quality that the ANN will use.

Reference [10] has compared several smoothing methods at the pre-processing stage. The compared methods were simple moving average (SMA), centered moving average (centered MA), double moving average (double MA), weighted moving average (weight MA) and modified weighted moving average algorithm (modified WMA). The predicted results show that centered MA smoothing can improve predictive accuracy performance. The previous study used the ES method to transform data before being used as input of the neural network [16]. The ES method was used to improve the quality of series data which would be used as training data for ANN. The ANN model was used to predict gold prices.

III. Data

The data used in this study were the amount of medicine used in Hospital "X". The two time series data showed the use of two drug types, Cephalexin (in milliliter) and Enxyplex (in grams). Time series data used was in the form of weekly data in 2015. Time series data for data 1 and data 2 can be seen in Fig. 1 and Fig. 2, respectively. The x-axis shows the week, and the y-axis indicates the number of medicines used.
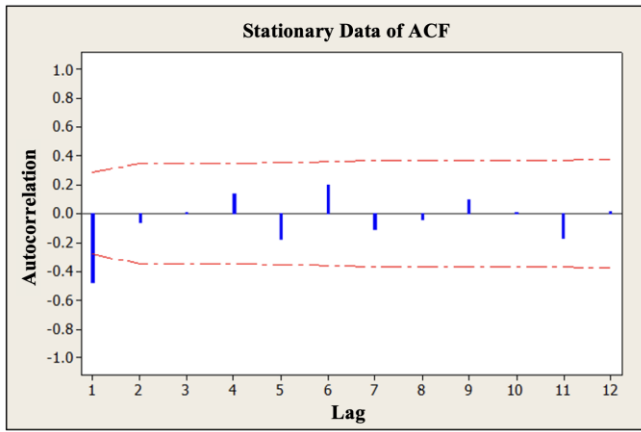
(a)
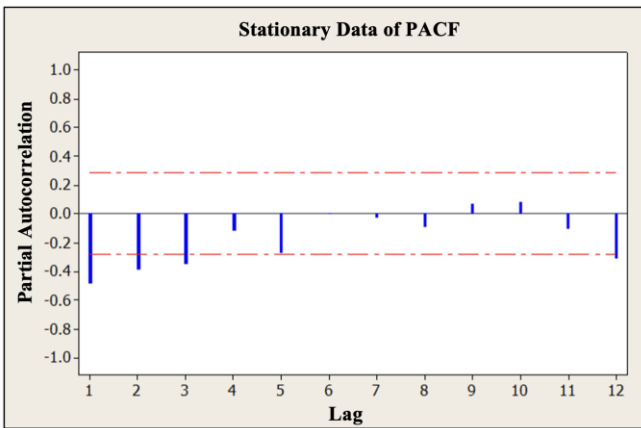
(b)

Fig. 3 Graph of stationary data, (a) ACF plot of stationary data, (b) PACF plot of stationary data.

IV. METHOD

The stages conducted in this study were reported as follows.

A. Lag Detection

In this stage, the time lag was detected for time series data using the autocorrelation function (ACF) and partial autocorrelation function (PACF). The dependencies between the actual data and the historical value represent the basic principle of the time series forecasting. The similarity of data for each value in a series of data adjacent to each other was observed to determine the dependencies. A statistical test to determine the dependencies between two entities is expressed as a correlation analysis to form a coefficient [12]. ACF can be used to analyze correlation in time series data. It detects a significant dependency on time series data [12].

In time series, the value of the first data is highly dependent on the value of the second data, the value of the second data is highly dependent on the value of the third data, and so on. It may result in the absence of significant dependency on the series. PACF can be used to solve the problem [12]. ACF and PACF analyses were used to determine the relevant input and the amount of lag utilized as inputs for ANN.

B. Time Series Data Transformation

In this stage, the time series data were normalized before being used as inputs of the ANN. Normalization is the step of adjusting data to a small range. In this study, the actual value was changed to the range of 0.1-0.9 using (1).

$$X' = \frac{0.8(x-a)}{b-a} + 0.1 \qquad (1)$$

where x is the value to be transformed, a is the smallest value of the time series data, and b is the largest value of the time series data.

C. Artificial Neural Network Structured Selection

The multilayer perceptron (MLP) network is one of popular neural network architectures. It can have one or more hidden layers between the input and output layers. The input node is a time lag, while the output node is the predicted value. The ANN model used in this study was the FFNN.

D. Whittaker-Henderson Smoothing Method

The WH method is one of the commonly used smoothing methods [23]. For n as the measurement sequence $\{y_1, y_2, ..., y_n\}$, a positive real value for λ, and a positive integer value $p < n$, then find the value $\{x_1, x_2, ..., x_n\}$ which minimizes (2).

$$\lambda \sum_{j=1}^{n}(y_j - x_j)^2 + \sum_{j=1}^{n-p}(\Delta^p x_j)^2 \qquad (2)$$

with the variable Δ is the forward difference operator.

For n as the measurement sequence $\{y_1, y_2, ..., y_n\}$, λ is the positive real value, and $p < n$ is the positive integer value. Then, the values of $\{x_1, x_2, ..., x_n\}$ that minimized the following equation ware calculated.

$$\Delta x_j = x_{j+1} - x_j \qquad (3)$$

$$\Delta^2 x_j = \Delta(\Delta x_j) = x_{j+2} - 2x_{j+1} + x_j \qquad (4)$$

with Δ is the forward difference operator.

In (3), the first sum was used to calculate the data fidelity, while the second sum (4) was used to calculate smoothness. In many cases, the p-value used is $p = 2$. The λ parameter controls the tradeoff between fidelity and smoothness [14]. When $\lambda \rightarrow 0$, then the estimated smoothness value is converged to p-1 of polynomial degree. Meanwhile, the estimated value is converged to the real data when $\lambda \rightarrow \infty$. The approach to select smoothness parameter λ is by trial and error or the GCV [15]. The selected λ was used to minimize the GCV score [14], [15], [23], [24].

$$GCV = \frac{1}{n}\sum_{j=1}^{n}\left(\frac{y_j - \bar{x}_j}{1 - n^{-1}trace(\lambda A^{-1})}\right)^2 \qquad (5)$$

with y is the real data, $\bar{x}$ is the data resulted from the smoothing process, and trace of (λA^{-1}) is also called the "hat" matrix.

V. EXPERIMENTAL RESULTS

A. Lag Detection

1) Dataset 1: Time series data needs to be stationary for looking at significant lags. The stationarity means the series stationary to mean and variance. The number of significant lags
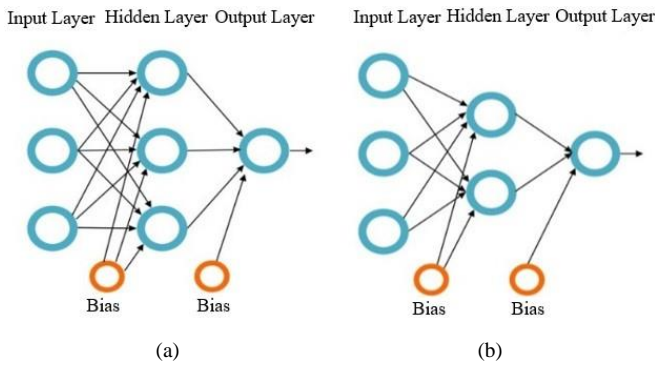
Fig. 4 ANN architecture for dataset 1, (a) the number of neurons in hidden layer = 3, (b) the number of neurons in hidden layer = 2.
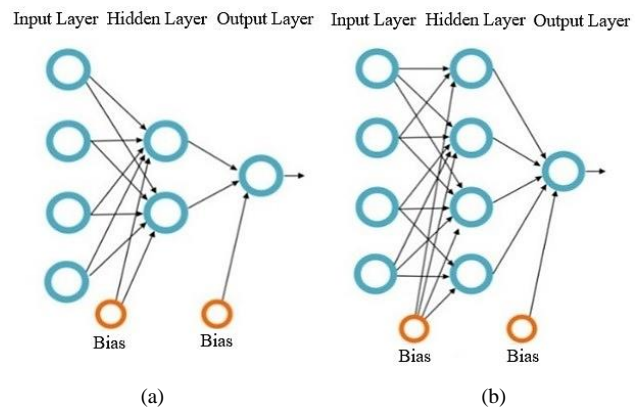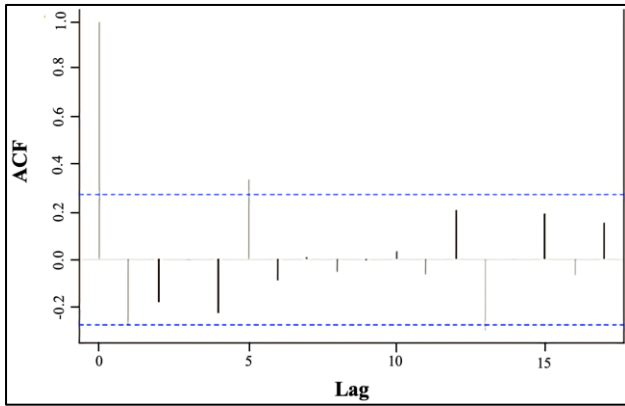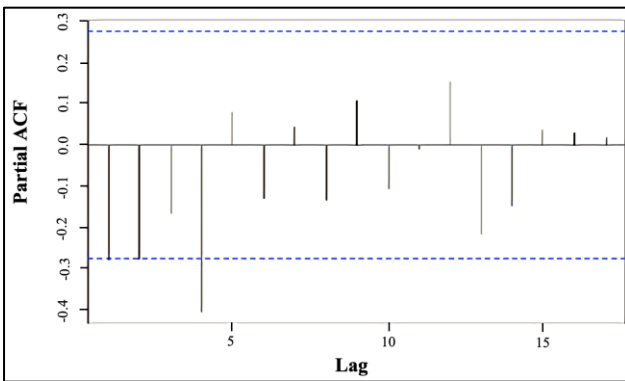

Fig. 6 ANN architecture used for dataset 2, (a) the number of neurons in hidden layer = 4, (b) the number of neurons in hidden layer = 2.
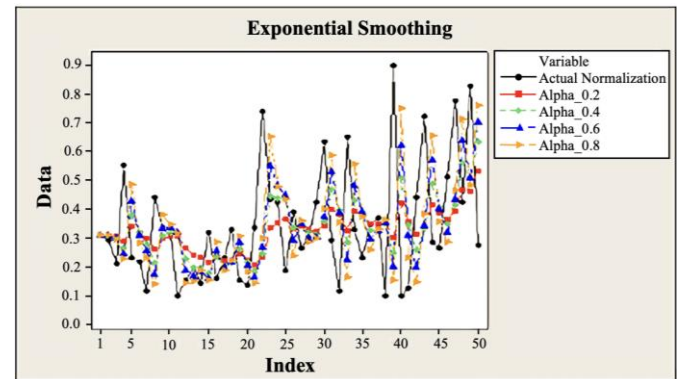

(a)

Fig. 7 Dataset 1 smoothing using simple ES.

ANN architectures provided were shown in Fig. 4 that shows the ANN architecture used for dataset 1 with the number of inputs neuron. Fig. 4(a) shows the number of neurons in hidden layer = 3 referring to Rule 1, while Fig. 4(b) shows the number of neurons in hidden layer = 2, referring to Rule 2.

2) Dataset 2: Fig. 5 shows lag detection results of dataset 2. Fig. 5(a) shows ACF plot stationary data which significant lags are lag 1 and lag 5. Meanwhile, Fig. 5(b), which is the PACF plot of stationary data, shows the significant lags are lag 1, lag 2, lag 4, and lag 5; then, the number of FFNN inputs are four input nodes.

The ANN architecture was adjusted by Rule 1 and Rule 2. In Rule 1, the number of neurons in the hidden layer was equal to the input, whereas in Rule 2, the number of neurons in the hidden layer was half the number of input neurons. The equations for Rule 1 and Rule 2 are shown in (2) and (3). They would provide two models of ANN architecture for dataset 2 with the number of inputs neuron = 4, shown in Fig. 6. Fig. 6(a) shows the number of neurons in hidden layer = 4 that refers to Rule 1, while Fig. 6(b) shows the number of neurons in hidden layer = 2 that refers to Rule 2.

(b)

Fig. 5 Lag detection results of dataset 2, (a) ACF plot stationary data and (b) PACF plot of stationary data.

of ACF and PACF will be used as the number of ANN input node. Fig. 3 shows lag detection results of dataset 1. ACF plot in Fig. 3(a) shows lag 1 is a significant lag and, in the PACF plot, Fig. 3(b) shows lag 1, lag 2, lag 3, and lag 12 are significant lags. Reference [25] suggests that a significant lag selection at the beginning of the ACF or PACF would be better; therefore, lag 1, lag 2, and lag 3 were selected to be the input of ANN. The ANN architecture was adjusted by Rule 1 and Rule 2. In Rule 1, the number of neurons in the hidden layer was equal to the input, whereas, in Rule 2, the number of neurons in the hidden layer was half the number of input neurons. The equations for Rule 1 and Rule 2 are shown in (6) and (7). Two

B. Time Series Data Transformation

1) Dataset 1: Before the transformation or smoothing stage was implemented, the time series data must be normalized first. In this stage, simple ES (Fig. 7) and WH smoothing (Fig. 8)
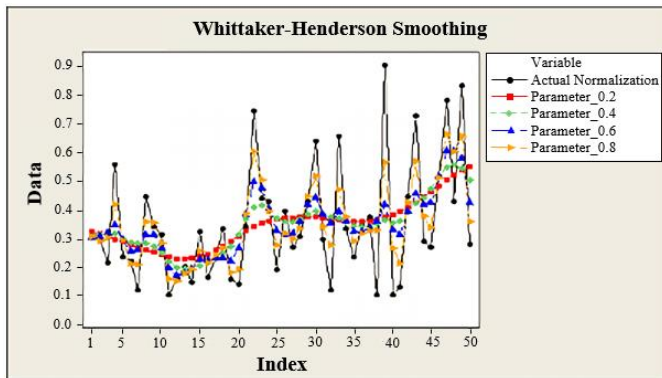
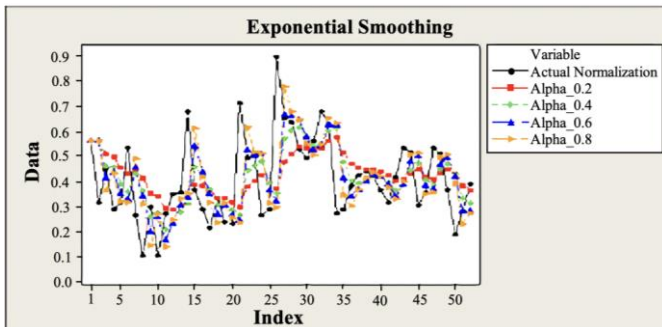Fig. 8 WH smoothing for dataset 1 with σ = 0.2; 0.4; 0.6; 0.8.

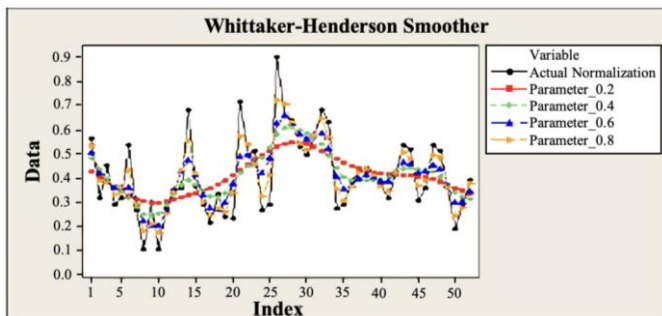Fig. 9 Dataset 2 smoothing using simple ES with σ = 0.2; 0.4; 0.6; 0.8.

Fig. 10 Dataset 2 smoothing using WH smoothing with σ = 0.2; 0.4; 0.6; 0.8.

were implemented to time series data then they would provide new time series dataset.

2) Dataset 2: Fig. 9 and Fig. 10 illustrate results of simple ES and WH smoothing into time series dataset 2.

C. Model Comparison

Determination of the number of ANN inputs used the ACF-PACF analysis. Tests were carried out on data with and without WH smoothing. The MSE results were then compared to find out the differences in the forecasting performance.

In this paper, there are two phases in the implementation of ANN. Phase 1 was conducted to find the best ANN structure by testing different learning rates (LR). After obtaining the best ANN structure, the best ANN structure was used for phase 2. In phase 2, it was tested against data that has been transformed using the ES method and WH. The learning algorithm used in this research was gradient descent with momentum (GDM). The learning parameters for the ANN-GDM algorithm are

TABLE I
ANN PARAMETER USED IN THIS RESEARCH

Parameter	Value
Epoch Maximum	1000
Targeted MSE	0
Maximum Validation Failure	6
Constant Momentum	0.7
Constant Learning Rate	0.1; 0.3; 0.5

TABLE II
THE BEST SMOOTHING PARAMETERS VALUE FOR EACH RULE OF THE
WH+ANN (WH+ANN) METHOD FOR THE DATASET 1

Rule	Learning Rate	Smoothing Parameter (σ)
Rule 1	0.3	0.1
Rule 1	0.5	0.1
Rule 2	0.5	0.1

TABLE III
THE BEST SMOOTHING PARAMETERS VALUE FOR EACH RULE OF THE
WH+ANN (WH+ANN) METHOD FOR THE DATASET 2

Rule	Learning Rate	Smoothing Parameter (σ)
Rule 1	0.3	0.8
Rule 1	0.5	0.7
Rule 2	0.5	0.9

listed in Table I. The best smoothing parameters values for dataset 1 and dataset 2 are presented in Table II and Table III.

The testing used one hidden layer and two conditions of the hidden neuron.
Rule 1:

$$Nh = Nin \qquad (6)$$

Rule 2:

$$Nh = Nin/2 \qquad (7)$$

with *Nh* = the number of hidden neurons; *Nin* = the number of ANN inputs.

Rule 1 was proposed by [26], where the effect of the number of hidden neurons on the performance of forecasting was investigated. In [27], several studies show that using neurons with a number equal to the number of input neurons generates good forecasting. Meanwhile, Rule 2 was proposed by [28], in which the number of hidden layer neurons was half the number of neurons in the network input. The activation function used in neurons in the hidden layer was hyperbolic tangent. The activation function used for output neurons was a linear activation function.

In this study, data sharing was divided into 70% for training, 15% for validation, and 15% for testing. The accuracy was measured by data testing. The forecasting error was measured using the MSE with the equation as follows.

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(y_t - y'_t)^2 \qquad (8)$$

with y_t is the actual value at time *t* and y'_t is the predicted value at time *t*. Tests were carried out 100 times with random initial weight in a small range of -1 to 1. An early stopping mechanism was used to prevent overfitting. The implementation of the ANN method used the Matlab software.

TABLE IV
DESCRIPTIVE STATISTIC OF SELECTED MODEL DATASET 1

Model	ANN			WH+ANN		
	Mean	Median	Std. Deviation	Mean MSE	Median MSE	Std. Deviation
Rule 1-LR 0.3	983.73	868.71	355.12	792.86	721.04	277.16
Rule 1-LR 0.5	949.80	830.05	630.78	768.26	733.31	142.79
Rule 2-LR 0.5	819.93	786.54	148.88	759.52	727.66	127.02

TABLE V
DESCRIPTIVE STATISTIC OF SELECTED MODEL DATASET 2

Model	ANN			WH+ANN		
	Mean	Median	Std. Deviation	Mean MSE	Median MSE	Std. Deviation
Rule 1-LR 0.3	256.79	182.80	375.65	219.41	178.50	245.62
Rule 1-LR 0.5	300.60	191.50	410.90	276.65	192.20	192.20
Rule 2-LR 0.5	217.37	188.00	153.48	222.08	185.20	275.95

After obtaining the best LR for each rule, the corresponding architecture was selected as the best architecture for the actual data. It was retested data smoothed using the WH (WH+ANN) method. The MSE results from the non-smoothed ANN model were compared with MSE from WH+ANN to determine whether there were differences in the predicted MSE value.

The test results based on the best LR parameter in the two rules for dataset 1 are shown in Table IV, while the results for dataset 2 are displayed in Table V. A statistical difference test between data groups was used to determine the best parameter value from ANN architecture. The MSE statistical test for both datasets shows that the best ANN model was obtained with LR of 0.3 and 0.5 for Rule 1; and LR of 0.5 in Rule 2. The LR of 0.3 and 0.5 in Rule 1 were selected because the test results for both LR show that the MSE acquisition for LR of 0.3 and 0.5 was not statistically different with a confidence level of less than 95%. Therefore, the two ANN models were used to represent the best model of Rule 1 and used for further testing. The results of the MSE acquisition from these models were then compared with the acquisition of MSE from the ANN with the WH method (WH+ANN).

The difference in the level of confidence in the MSE value of the WH+ANN method in the dataset 2 could occur due to the differences of spikes pattern between the two datasets. The observation of the time series data plots shows that the dataset 2 containing spikes or fluctuating patterns is fewer than the dataset 1, indicating that the smoothing process for the dataset 2 does not give a significant smoothing effect. Therefore, the comparison results of MSE between WH+ANN and ANN for dataset 2 did not produce a statistically significant difference in a significance level of 0.05.

The test results for dataset 1 can be seen in Table IV, which shows that the MSE acquisition from WH+ANN is smaller than the acquisition of MSE from ANN for both Rule 1 and Rule 2. Meanwhile, Table V displays the test results in dataset 2. As it can be seen in Table V, the MSE acquisition from WH+ANN tends to have a smaller value than the MSE acquisition from ANN for Rule 1 and Rule 2. The results demonstrate that the ANN method that used the WH (WH+ANN) smoothing method had a significantly lower MSE than the non-smoothing ANN method with a confidence level of 94% for dataset 1 and 85% for dataset 2.

The experiment results show that proper selection of smoothing parameters value may produce low MSE values. However, if the selection of smoothing parameters value is incorrect, it may produce a high MSE. Determining an appropriate value of smoothing parameters is a challenge encountered in the application of smoothing methods in the pre-processing phase of time series data for improving the accuracy of ANN forecasting. The results indicate that the WH smoothing parameter can produce the lowest MSE if it has the smallest GCV value. However, it is necessary to investigate whether low GCV always produces a low MSE forecasting for ANN.

VI. CONCLUSIONS

This study focused on the application of smoothing methods at the pre-processing stage of the ANN method. In significant lags, time series data needs to be stationary. The stationary condition suggests the series are stationary to mean and variance. The number of significant lags of ACF and PACF would be used as the number of ANN input nodes.

The comparative predictive analysis was done by comparing the ANN method with and without smoothing. The smoothing method used in this study was the WH. The results show that WH+ANN has a statistically lower MSE than ANN (without smoothing), with a confidence level of 94% for dataset 1 and 85% for dataset 2. It indicates that the data smoothing process affects the accuracy of the ANN model forecasting.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTION

REFERENCES

[1] M.I. Ramos, J.J. Cubillas, and F.R. Feito, "Improvement of the Prediction of Drugs Demand Using Spatial Data Mining Tools," *J. Med. Syst.*, Vol. 40, No. 1, pp. 1-9, Jan. 2016.

[2] H. Pratyaksa, A.E. Permanasari, S. Fauziati, and I. Fitriana, "ARIMA Implementation to Predict the Amount of Antiseptic Medicine Usage in Veterinary Hospital," *2016 1st Int. Conf. Biomed. Eng. (IBIOMED)*, 2016, pp. 1-4.

[3] C. Qingkui and R. Junhu, "Study on the Demand Forecasting of Hospital Stocks Based on Data Mining and BP Neural Networks," *2009 Int. Conf. Electron. Commerce, Bus. Intell.*, 2009, pp. 284-289.

[4] M. Jalalpour, Y. Gel, and S. Levin, "Forecasting Demand for Health Services: Development of A Publicly Available Toolbox," *Oper. Res. Health Care*, Vol. 5, pp. 1-9, Jun. 2015.

[5] N.K. Zadeh, M.M. Sepehri, and H. Farvaresh, "Intelligent Sales Prediction for Pharmaceutical Distribution Companies: A Data Mining Based Approach," *Math. Probl. Eng.*, Vol. 2014, pp. 1-15, May 2014.

[6] Á. Lublóy, "Factors Affecting the Uptake of New Medicines: A Systematic Literature Review," *BMC Health Serv. Res.*, Vol. 14, pp. 1-25, Oct. 2014.

[7] D. Alba-Cuéllar, *et al.*, "Time Series Forecasting with PSO-Optimized Neural Networks," *2014 13th Mexican Int. Conf. Artif. Intell.*, 2014, pp. 102-111.

[8] F. Yu and X. Xu, "A Short-Term Load Forecasting Model of Natural Gas Based on Optimized Genetic Algorithm and Improved BP Neural Network," *Appl. Energy*, Vol. 134, pp. 102-113, Dec. 2014.

[9] I. Khandelwal, R. Adhikari, and G. Verma, "Time Series Forecasting Using Hybrid ARIMA and ANN Models Based on DWT Decomposition," *Procedia Comput. Sci.*, Vol. 48, pp. 173-179, 2015.

[10] F. Nhita, D. Saepudin, Adiwijaya, and U.N. Wisesty, "Comparative Study of Moving Average on Rainfall Time Series Data for Rainfall Forecasting Based on Evolving Neural Network Classifier," *2015 3rd Int. Symp. Comput., Bus. Intell. (ISCBI)*, 2015, pp. 112-116.

[11] D.I. Wilson, "The Black Art of Smoothing," *Elect., Automat. Technol.*, June/July Issue, pp. 35-36, 2006.

[12] O. Ostashchuk, "Time Series Data Prediction and Analysis," Master thesis, Czech Technical University in Prague, Prague, Czech Republic, 2017.

[13] D. Garcia, "Robust Smoothing of Gridded Data in One and Higher Dimensions with Missing Values," *Comput. Statist., Data Anal.*, Vol. 54, No. 4, pp. 1167-1178, Apr. 2010.

[14] H.L. Weinert, "Efficient Computation for Whittaker-Henderson Smoothing," *Computat. Statist., Data Anal.*, Vol. 52, No. 2, pp. 959-974, Oct. 2007.

[15] J.J. Stickel, "Data Smoothing and Numerical Differentiation by A Regularization Method," *Comput., Chem. Eng.*, Vol. 34, No. 4, pp. 467-475, Apr. 2010.

[16] I. Suryani and R.S. Wahono, "Penerapan Exponential Smoothing untuk Transformasi Data dalam Meningkatkan Akurasi Neural Network pada Prediksi Harga Emas," *J. Intell. Syst.*, Vol. 1, No. 2, pp. 67-75, Dec. 2015.

[17] I.N. Soyiri and D.D. Reidpath, "An Overview of Health Forecasting," *Environ. Health, Prev. Med.*, Vol. 18, pp. 1-9, Jan. 2013.

[18] I.K. Utami, "Seleksi Input untuk Artificial Neural Network Menggunakan Binary Particle Swarm Optimization dalam Pemodelan Runtun Waktu Kasus Avian Influenza," Master thesis, Universitas Gadjah Mada, Yogyakarta, Indonesia, 2016.

[19] G. Lachtermacher and J.D. Fuller, "Backpropagation in Time-Series Forecasting," *J. Forecast.*, Vol. 14, No. 4, pp. 381-393, Jul. 1995.

[20] F.S. Wong, "Time Series Forecasting Using Back Neural Networks," *Neurocomputing*, Vol. 2, No. 4, pp. 147-159, Jul. 1991.

[21] A. Azadeh, M. Sheikhalishahi, M. Tabesh, and A. Negahban, "The Effects of Pre-Processing Methods on Forecasting Improvement of Artificial Neural Networks," *Aust. J. Basic, Appl. Sci.*, Vol. 5, No. 6, pp. 570-580, Jun. 2011.

[22] S. Anbazhagan and N. Kumarappan, "Day-Ahead Deregulated Electricity Market Price Forecasting Using Neural Network Input Featured by DCT," *Energy Convers., Manage.*, Vol. 78, No. 2, pp. 711-719, Feb. 2014.

[23] A.S. Nocon and W.F. Scott, "An Extension of the Whittaker–Henderson Method of Graduation," *Scand. Actuar. J.*, Vol. 2012, No. 1, pp. 70-79, Mar. 2012.

[24] P.H.C. Eilers, "A Perfect Smoother," *Anal. Chem.*, Vol. 75, No. 14, pp. 3631-3636, May 2003.

[25] H. Abolfazli, S.M. Asadzadeh, and S.M. Asadzadeh, "Forecasting Rail Transport Petroleum Consumption Using an Integrated Model of Autocorrelation Functions-Artificial Neural Network," *Acta Polytech. Hungarica*, Vol. 11, No. 2, pp. 203-214, Jan. 2014.

[26] Z. Tang and P.A. Fishwick, "Feedforward Neural Nets as Models for Time Series Forecasting," *ORSA J. Comput.*, Vol. 5, No. 4, pp. 374-385, Nov. 1993.

[27] P.G. Zhang, E. Patuwo, and M. Y. Hu, "Forecasting with Artificial Neural Networks: The State of the Art," *Int. J. Forecasting*, Vol. 14, No. 1, pp. 35-62, Mar. 1998.

[28] S.Y. Kang, "An Investigation of the Use of Feedforward Neural Networks for Forecasting," Ph.D. dissertation, Kent State University, Ohio, USA, 1991.