Recurrent Neural Network For Forecasting Time Series With Long Memory Pattern

by Walid Matematika

Journal of Physics: Conference Series

PAPER • OPEN ACCESS

Recurrent Neural Network For Forecasting Time Series With Long Memory Pattern

Recurrent Neural Network For Forecasting Time Series With Long Memory Pattern

Walid and Alamsyah

Department of Mathematics, Universitas Negeri Semarang

Corresponding author: alikuwalid87@yahoo.com

Abstract. Recurrent Neural Network as one of the hybrid models are often used to predict and estimate the issues related to electricity, can be used to describe the cause of the swelling of electrical load which experienced by PLN. In this research will be developed RNN forecasting procedures at the time series with long memory patterns. Considering the application is the national electrical load which of course has a different trend with the condition of the electrical load in any country. This research produces the algorithm of time series forecasting which has long memory pattern using E-RNN after this referred to the algorithm of integrated fractional recurrent neural networks (FIRNN).The prediction results of long memory time series using models Fractional Integrated Recurrent Neural Network (FIRNN) showed that the model with the selection of data difference in the range of [-1,1] and the model of Fractional Integrated Recurrent Neural Network (FIRNN) (24,6,1) provides the smallest MSE value, which is 0.00149684.

1. Introduction

Electricity plays a significant role in realising the strategic objectives and national development, and therefore the power supply business controlled by the state and provision should be increased in line with the elaboration of the construction that provided electricity in sufficient quantities, equitable and quality. Supply of electric power is capital intensive and technology and in line with the principle of decentralisation and democratisation in the order of society, nation, and state the role of local governments and communities in the provision of electric power needs to be improved.

The importance of energy for society can be indicated by the amount of electricity used by the public both for household consumption and for industrial and business activity in local and national scale. Obviously, this will significantly affect the quantity and quality of goods or services that exist. Another sector that is not less necessary for the functioning of electricity are the advances in communications technology and informatics who took part in expanding the space for the transportation of goods and services in all areas of life today.

According to Law No. 30 of 2009 on electricity, PT. PLN (State Electricity Company) as the holder of the power of electricity business is required to provide continuous electrical power (constant) with high quality and reliability, are also needed to provide excellent service to customers/consumers of electricity. It turns out the circumstances encountered much different now to what has been established by the Electricity Act. As a result of the power supply crisis experienced in almost all corners of the country, PT. PLN is not able to provide continuous power to its customers.

Estimate or forecast the amount of electricity used to optimise the use of power in society, so that there will be no wastage or power outage. According to F. Chan and C. Lim (2011) forecasting is an important tool in planning the effective and efficient.So if it based on a count of the Central Bureau of Statistics which is abbreviated by BPS, there are 50 billion SMEs, mean total losses reached 2.5 billion unavoidable due to blackouts by PLN. It is not surprising because the electrical energy is indeed a prime mover in industrial activity and a full range of human needs. Electricity has become a necessity indispensable to man as closely linked to the overall economic base. This ranging from the smallest to the greatest extent.Specifically, in urban areas, dependence on electrical energy is very high, and it concerns the core activities within the scope of the industry that allows subdued overall economic security.

Other aspects that also suffered a setback as a result of planned blackouts by PLN is the effect on the decrease in the interest of investors to invest in Indonesia because he felt there was no guarantee of the power supply becomes the driving force of industrial production processes. According to [1], there are several stages in forecasting, among others, model identification, parameter estimation, verification of models and forecasting. At the time series prediction models forecasting the future state based on the past. Time series approach can use several methods, may include analysis using autocorrelation function and partial function autocorrelation to study changes with time series parametric models known as ARIMA time domain analysis.

Besides, there is an alternative approach that used as a method of spectral analysis. The spectral method is one form of the Fourier transform.To be able to forecast correctly then one of the first steps is to identify the data model to know whether the data included in the data group which has a long memory categories or short memory. Identification of this initial data will provide valuable information relating to the next step of estimating the parameters towards-existing parameters, then verify and new forecasting. Information good start would be a significant contribution in providing a level of accuracy of prediction.

Based on the above background, this paper will focus on a study to identify the model of extended memory using the rescaled range. The data is the data used to simulate the electrical load is calculated every hour throughout the year. Issues to be studied is how does the identification of long memory models using the rescale range?

2. Time Series Analysis And Long Memory Model

Time series analysis by [2] has several purposes, namely forecasting, modelling, and control. Forecasting issues related to the establishment of models and methods that can be used to produce an accurate estimate. Modelling aims to obtain an appropriate statistical model to represent the behaviour of a long-term time series data. Difference modelling with forecasting is predicting more likely on a model of the "black-box" to get the forecast, while modelling tends to models that can be interpreted to explain what is going on concerning the relationship between variables in a time series data. While the purpose of the control widely used in engineering, especially signal processing.

Statistical modelling of time series analysis if traced back, preceded by Yule (1927) which introduced autoregressive linear models (AR) to predict yearly sunspot numbers. Since the publication concerned with the analysis of time series is growing rapidly. Until 1980, most of the research has focused on the model of linear time series, mainly linear model class Autoregressive Integrated Moving Average (ARIMA). [1] developed a complete procedure for ARIMA model methodology hitherto used as a standard procedure in the establishment linear time series models. Some literature which discusses many ARIMA model seen in [3] and [4]. Also, the properties related to statistical theory for the ARIMA model has also been extensively analysed and developed by several researchers, among others, have been conducted by [5].

Time series is a series of observations Y_t on a variable Y, which each observation recorded at a particular time $t \in T$. In this case, T is the set of time in which the comments made. If T is a discrete set, then $\{Y_t,\ t \in T\}$ is a separate time series. $\{Y_t,\ t \in T\}$ is a notation whole time series, where Yt is the observation of $\{Y_t,\ t \in T\}$ in time to t. In the case of discrete time series, views usually taken at the same time interval.

A statistical approach to the analysis of time series conducted using statistical models to describe the dynamic behaviour of a time series. It is assumed that a time series generated by a mechanism or a stochastic model, which is usually defined by a stochastic difference equation. It is stochastic difference equation that consists of an equation and some initial conditions. Results or a solution of this model is a stochastic process, which is a sequence of random variables $\{Y_t\}$ defined on the probability space (Ω, $\mathfrak{K}$, P). For certain ω□Ω, Y_t(ω) is called a realisation (sample path or trajectory) of $\{Y_t\}$ Each observation

$Y_t(\omega)$ is a realisable value of the random variable Y_t whose values obtained in d-dimensional Euclidean space $\mathbb{R}^d$.

The most significant assumptions in the analysis of time series are stationary data. Nonstationarity inspection data done with the help of time series plots using scatter plots. Nonstationarity in the mean can be resolved by the process of differentiating (differencing). According to [1] stationary data invariance can be seen with a value of λ (lambda) estimate the Box-Cox transformation

[5] expressed a process $\{Y_t\}$ is said to be stationary stronger if $(y_{t_1}, y_{t_2}, \ldots, y_{t_k})'$ and $(y_{t_{1+l}}, y_{t_{2+l}}, \ldots, y_{t_{k+1}})'$ have the same joint distribution function for all integers k≥1dan for all $t_1, t_2, \ldots, t_k, k \in \mathbb{N}$. Processes with the first and second moments that are independent of time is also a concern in time series analysis. The following definitions relating to the concept of a weak fixed or stationary until the second order.

If given $\{Y_t\}$ is a process with $E|Y_t|^2 < \infty$ for every $t \in \mathbb{N}$, then $\{Y_t\}$ is said to be fixed weak (weakly stationary) if $E(Y_t) = \mu$ for all $t \in \mathbb{N}$ and $\mathrm{Cov}(Y_r, Y_s) = \mathrm{Cov}(Y_{r+h}, Y_{s+h})$ for all $r, s, h \in \mathbb{N}$. If $\{Y_t\}$ is a process that is stationary weak (weakly stationary), then the function autocovariance $\gamma_y(.)$ of $\{Y_t\}$ defined $\gamma_y(h) = \mathrm{Cov}(Y_t, Y_{t+h})$ to all $t, h \in \mathbb{N}$ [5].

Time series that have a long-term dependence, or better known as the long memory is a time series in which observations are far apart still have a high correlation. Autocorrelation of a process of long memory fell hyperbolic. A sequence $\{Y_t\}$ is said to follow the model Autoregressive Integrated Moving Average if the distinction to d namely $W_t = \nabla^d Y_t$ is a stationary ARMA process. If W_t is ARMA (p, q), then Y_t is ARIMA (p, d, q). In practice, the value of d used is 1 or at most 3 [3]. Model ARFIMA able to model the dependence of short-term and long-term. Observations generated by the structure of ARMA shows the relationship of short, while the distinction fractional parameter d, which caused values to fall hyperbolic ACF showed long-term dependence.

According to Wilfredo Palma (2007), a definition of long memory understood as follows. Let $\gamma(h) = \langle y_t, y_{t+h} \rangle$ is the covariance function of the lag h of a stationary process $\{y_t : t \epsilon \mathbb{Z}\}$. The usual definition of long memory is

$$\sum_{h=-\infty}^{\infty} |\gamma\ (h)| = \infty \tag{1}$$

However, there is an alternative explanation. In particular, long memory can be defined by specifying the hyperbolic decline in this autocovariance. Recalling that a function of measuring active defined in some environments [a, ∞) infinite is said to change slowly in a sense because if and only if for every $c > 0, l(cx)/l(x)$ converges to 1 of x tends Infinity. Examples slowly changing function is $l(x) = \log(x)$ and $l(x) = b$, where b is a positive constant. Furthermore, notation $x_n \sim y_n$ which means that $\frac{x_n}{y_n} \to 1$ of $n \to \infty$, unless otherwise stated.

$$\gamma\ (h) \sim h^{2d-1} l_1(h) \tag{2}$$

For $h \to \infty$, where d is called long memory parameters and $l_1(.)$ is a function that is changing slowly.

Another widely used definition of a high dependence on the spectral domain is

$$f(\lambda) \sim |\lambda|^{-2d} l_2(1/|\lambda|) \tag{3}$$

To λ in zero environments and $l_2(.)$ is a slowly changing function. Furthermore, an alternative definition of long memory on the rules that are based directly on the decomposition Wold

$$\psi_j \sim j^{d-1} l_3(j) \tag{4}$$

For j> 0, where $l_3(.)$ is a slowly changing function. Model ARFIMA (p, d, q) developed [6] is written as:

$$\phi(B)(1 - B)^d (Y_t - \mu) = \theta(B) a_t \tag{5}$$

With t: index of observation

d: distinguishing parameters (fractions)

IOP Conf. Series: Journal of Physics: Conf. Series **824** (2017) 012038 doi:10.1088/1742-6596/824/1/012038

μ: average observation

$a_t \sim IIDN\ (0, \sigma_a^2)$,

$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$ is a polynomial AR (p),

$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$ q are polynomials MA (q),

$(1-B)^d = \nabla^d = \sum_{k=0}^{\infty} \binom{d}{k} (-1)^k B^k$ differentiator fractional operator,

$$\sum_{k=0}^{\infty} \binom{d}{k} (-1)^k B^k = F(-d, 1; 1; B)$$

$$= \sum_{k=0}^{\infty} \frac{\Gamma(k-d)}{\Gamma(k+1)\Gamma(-d)} B^k$$

F is the hypergeometric function is defined as follows.

$$F(a, b; c; d) = 1 + \frac{a.b}{1.c} B + \frac{a(a+1)b(b+1)}{1.2.c(c+1)} B^2 + \cdots \quad (6)$$

a, b and c are real numbers and B is the operator Backshift.

For d is value fraction, fractional differencing operator $(1-B)^d$

Defined as

$$(1-B)^d = 1 + \sum_{k=1}^{\infty} \frac{\Gamma(-d+k)}{\Gamma(-d)k!} B^k \quad (7)$$

If the equation $\lambda_k(d) = \frac{\Gamma(-d+k)}{\Gamma(-d)k!}$ in equation (7) described for various values of k then:

For k = 1, is obtained $\frac{\Gamma(-d+1)}{\Gamma(-d)1!} = \frac{(-d)!}{(-d-1)!1!} = -d$,

For k=2, diperoleh $\frac{\Gamma(-d+2)}{\Gamma(-d)2!} = \frac{(-d+1)!}{(-d-1)!2!} = \frac{-d(1-d)}{2}$,

k=3, diperoleh $\frac{\Gamma(-d+3)}{\Gamma(-d)3!} = \frac{(-d+2)!}{(-d-1)!3!} = \frac{-d(1-d)(2-d)}{6}$ and so on so that the equation (3.3) can be written back into

$$(1-B)^d = 1 + \sum_{k=1}^{\infty} \lambda_k B^k \quad (8)$$

with $\lambda_0(d) = 1, \lambda_1(d) = -d$ $\lambda_2(d) = -\frac{1}{2}d(1-d), \lambda_3(d) = -\frac{1}{6}d(1-d)(2-d)$ and so on. So that equation (8) above can be written back into

$$(1-B)^d = 1 - dB - \tfrac{1}{2}d(1-d)B^2 - \tfrac{1}{6}d(1-d)(2-d)B^3 - \cdots \quad (9)$$

In general model of seasonal autoregressive integrated moving average (ARIMA) (P, D, Q) S is as follows.

$$\Phi_p(B^S)(1-B^S)^D Z_t = \Theta_Q(B^S) a_t \quad (10)$$

with $\Phi_p(B^S) = 1 - \Phi_1 B^S - \Phi_2 B^{2S} \ldots - \Phi_P B^{PS}$

$$\Theta_Q(B^S) = 1 - \Theta_1 B^S - \Theta_2 B^{2S} \ldots - \Theta_Q B^{QS}$$

Multiplicative ARIMA model written in the form

$$\phi_p(B)\Phi_p(B^S)(1-B)^d(1-B^S)^D Z_t = \theta_q(B)\Theta_Q(B^S) a_t \quad (11)$$

This model is often called by the model SARIMA (p, d, q) (P, D, Q)^{S}S

This model reduces to the ARIMA model (p, d, q) when no seasonal effects as well as being ARMA (p, q) when the stationary time series.

Additive SARIMA generalized models can be written in the form

$$\left(1-\phi_1 B-\phi_2 B^2-\cdots-\phi_p B^p-\Phi_1 B^S-\Phi_2 B^{2S}\ldots-\Phi_P B^{PS}\right)(1-B)^d(1-B^S)^D Z_t$$
$$=\left(1-\theta_1 B-\theta_2 B^2-\cdots-\theta_q B^q-\Theta_1 B^S-\Theta_2 B^{2S}\ldots-\Theta_Q B^{QS}\right)a_t \qquad (12)$$

3. Research Methods

The data used in this study is the simulation data generated from the values of Hurst, including Hurst value: H = 0.12345; H = 0.1567; H = 0.2389; H = 0.3456; H = 0.4535; H =0.4995; H = 0.5000; H = 0.5235; H = 0.5853; H = 0.6534; H = 0.7654; H = 0.8745; H = 0.90345; H = 0.9999.

In addition to the above simulation data, this study also uses data the electricity consumption data which is secondary data from the PLN in Java and Bali. Data is data consumption every half-hour heat load Se Bali-Java, taken from January 1 to the date of the end of December 2010. Through the program Minitab and Matlab searchable form plot (scatter) data. Furthermore, the data in the analysis to determine the magnitude of the value difference d (difference) method rescale range (R/S). This process begins with determining the average value of the data, the adjusted mean and standard deviation.

Furthermore, the determination of the value of the cumulative deviation and range of the cumulative difference, and the final stage is to determine the statistical value of Hurst (H) with logarithmic statistics (R/S) and assess the value of H through Ordinary Least Square method (OLS).

4. Result

Results of analysis with Matlab program for Hurst value seen in figure 1.

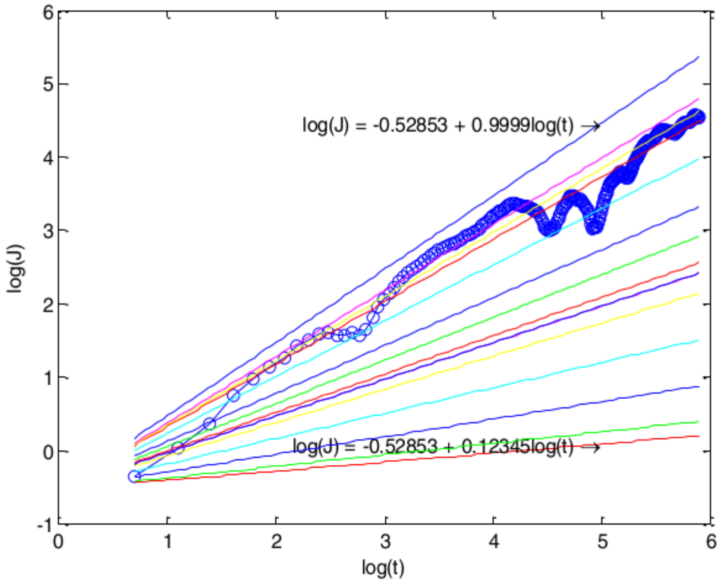

Figure 1 Hurst Value

Preprocessing done by using transformation, in this study conducted transformations on the data difference in the range [1,1] and conversion of data difference in the range [-1,1]. While the first conversion function used is function prestd (targtrain) to obtain the target data normally distributed, namely the achievement of a zero mean value and variance value is one. While the role of the second transformation in the preprocessing is by using the conversion prestd (inptrain) with the aim to obtain

IOP Conf. Series: Journal of Physics: Conf. Series **824** (2017) 012038 doi:10.1088/1742-6596/824/1/012038

the input data were normally distributed, namely the achievement of a zero mean value and variance value is one. The result of the preprocessing stage using data difference through two models namely the type function net = newff (minmax (in), [1,1], { 'tansig', 'purelin'}) and the model function of net = train (net, in, tn).

The next step is to provide training using difference NN data on the range [1,1], the two models in NN used that model RNN. Training in RNN model by modifying the number of hidden layers to obtain optimal results in forecasting, optimisation methods used in training are Levenberg Marquardt. The optimal lead to forecasting this will be demonstrated by evaluating the value of MSE on each model of training used.

Based on the above steps, obtained several models for use in forecasting time series patterned long memory. The results of model identification, among others model FIRNN (24,1,1), FIRNN (24,2,1), FIRNN (24,3,1), FIRNN (24,4,1), FIRNN (24,5,1), FIRNN (24,6,1), and model FIRNN (24, 7.1).

Also, the results of training iterations difference 24 RNN with the data in the range [1,1] each model found in the appendix. FIRNN training results using data difference in the range [-1,1]. Some models RNN also developed, with the type identification obtained several models. Among others: the model RNN (24,1,1) using 1 unit hidden and 1 unit of output and optimisation methods used in training are Levenberg Marquardt, the results obtained in the epoch to 21 model provides MSE value of 0.00570881 with an amount of gradient 0,23376x10^{-5}. It appears that at this period to 23 models reach towards convergence with absolute values. The results of the iteration shown in figure 2.

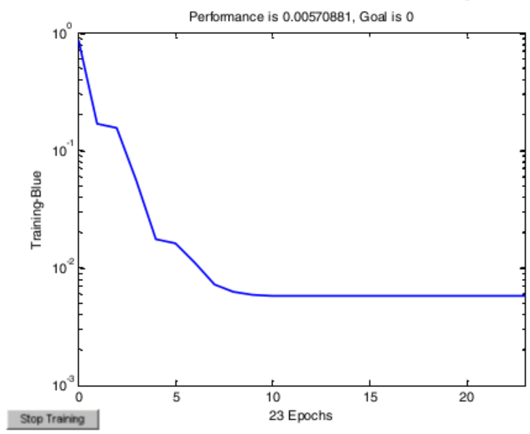

Figure 2 The Results Of Iteration

Training with the model FIRNN (24,2,1) using 2 hidden units and 1 unit of output and optimisation methods used in training are Levenberg Marquardt, the results obtained in the epoch to the 100 model provides MSE value to the value of 0.0392602 with gradient 2, 11941x10^{-5}. This model produces period to 100 to reach convergent. The results of the iteration shown in figure 3.

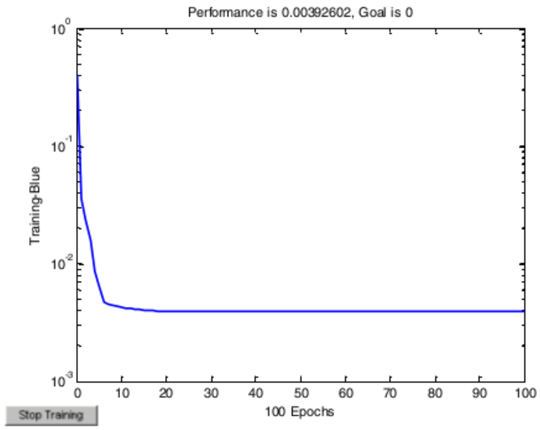

Figure 3 The Results Of Iteration

Training models FIRNN (24,3,1) with 3 hidden units and 1 unit of output and optimisation methods used in training are Levenberg Marquardt, the performance in the epoch of the 100 model provides MSE value of 0.0301456 with gradient value 0, 48871x10^{-5}. This model produces period to 100 to reach convergent. The results of the iteration shown in figure 4.

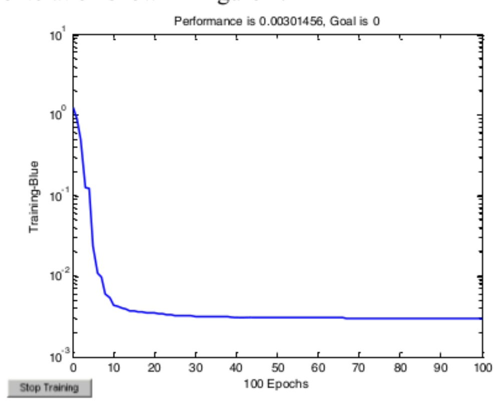

Figure 4 The Results Of Iteration

Training models FIRNN (24,4,1) using 4 units hidden and 1 unit of output and optimisation methods used in training are Levenberg Marquardt; the result converges on epoch to 100, this model provides MSE value of 0.00240606 and value of gradient by 1,31178x10^{-5}. Training models FIRNN (24,5,1) using 5 hidden units and 1 unit of output and optimisation methods used in training are Levenberg Marquardt; the result converges on epoch to 100, this model provides MSE value of 0.00202306 value gradient 0,34653x10^{-5}. Training models FIRNN (24,6,1) using 6 hidden units and 1 unit of output and optimisation methods used in training are Levenberg Marquardt, obtained results converge to the epoch to 100 model provides MSE value of 0.00177697 with value amounted gradient 0,37509x10^{-5}.

Training models FIRNN (24,7,1) using 7 hidden units and 1 unit of output and optimisation methods used in training are Levenberg Marquardt, obtained results converge to the epoch to 100 model provides MSE value of 0.00170185 and value gradient 0,54 130 x10^{-5}.

5. Conclusions

The results of long memory time series forecasting using models Fractional Integrated Recurrent Neural Network (FIRNN) showed that the model Fractional Integrated Recurrent Neural Network (FIRNN) (24,7,1) gives the smallest MSE value, which is 0.00177697. Besides, using simulated data, the result that for the simulated data with H <0.5 forecasting results obtained by using FIRNN shows the level of accuracy of 80%, while for H> 0.5 provides a 90% accuracy rate. Use of the activation function in the E-RNN needs to be modified with other functions to get the variance results in line with expectations so that it will be obtained an activation function appropriately in forecasting.

References

[1] Box G E P & Jenkins G M 1976 *Time Series Analysis, Forecasting and Control, 3rd Edition* (Englewood Cliffs: Prentice Hall)
[2] Chatfield C 2001 *Time Series Forecasting* (Chapman)
[3] Wei W W S 1999 *Time Series Analysis: Univariate and Multivariate Methods* (USA: Addison-Wesley Publishing Co)
[4] Box E.P George *et al*. 1994 *Time Series Analysis, Forecasting and Contro* (New Jersey: Prentice Hall Inc)
[5] Brockwell P J and Davis RA 1991 *Time Series Theory and Methods, 2nd Edition* (New York: Springer-Verlag)
[6] Joyeuk 1980 An Introduction to Long Memory Time Series Models and Fractional Differencing, Journal of Time Series Analysis, Vol. 1, pp. 15-29.

Recurrent Neural Network For Forecasting Time Series With Long Memory Pattern

ORIGINALITY REPORT

17%	14%	12%	11%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

MATCH ALL SOURCES (ONLY SELECTED SOURCE PRINTED)

7%

★ Suhartono, S.P. Rahayu, D.D. Prastyo, D.G.P. Wijayanti, Juliyanto. "Hybrid model for forecasting time series with trend, seasonal and salendar variation patterns", Journal of Physics: Conference Series, 2017

Publication

Exclude quotes	On	Exclude matches	< 15 words
Exclude bibliography	On		