



Modelling Multi Input Transfer Function for Rainfall Forecasting in Batu City

Priska Arindya Purnama, Ni Wayan Surya Wardhani, Rahma Fitriani

Department of Statistics, Brawijaya University, Malang

Email: priska.arindya@gmail.com, wswardhani@yahoo.com, rahmafitriani@ub.ac.id

ABSTRACT

The aim of this research is to obtain the appropriate transfer function to model rainfall data in Batu City and to figure out how accurate the results of rainfall forecasting in Batu City using the transfer function model based on air temperature, humidity, wind speed and cloud. Transfer function model is a multivariate time series model which consists of an output series (Y_t) sequence expected to be affected by an input series (X_t) and other inputs in a group called a noise series (N_t). Air temperature, humidity, wind speed and cloud are used as an input series (X_t) and rainfall as an output series (Y_t). Multi input transfer function model obtained is $(b_1, s_1, r_1) (b_2, s_2, r_2) (b_3, s_3, r_3) (b_4, s_4, r_4) (p_n, q_n) = (0,0,0) (23,0,0) (1,2,0) (0,0,0) ([5,8],2)$ and shows that rainfall on a certain day is affected by air temperature and cloud on that day, air humidity in the previous 23 days and wind speed in the previous day. The results of rainfall forecasting in Batu City using multi input with single output transfer function model is accurate based on the result of model validation using t test and MAPE statistic that less than 20%.

Keywords: multi input transfer function model, rainfall, forecasting

INTRODUCTION

Time series analysis is a data analysis considering the effect of time. Time series or time series data is a sequence of observations taken sequentially based on time with the same interval, ie daily, weekly, monthly, yearly or other time periods [1]. The modeling in time series analysis is classified into two, ie univariate and multivariate. One univariate time series model is Autoregressive Integrated Moving Average (ARIMA), whereas one multivariate time series model is a transfer function model. The transfer function model consists of an output series (Y_t) that is influenced by the input series (X_t) and other inputs combined in a group called the noise series (N_t) [2]. The transfer function model combines several characteristics of the univariate ARIMA model and multiple regression analysis or combines time series analysis with a causal approach. The purpose of modeling the transfer function is to establish a simple model that can connect the output series (Y_t) with the input series (X_t) and the noise series (N_t).

Some forecasting research using transfer function model is done by Tankersley, et. al. [3] about groundwater fluctuations in Florida showed that the transfer function model is more accurate than ARIMA. While Edlurd and Karlsson [4] who studied the unemployment rate in Sweden showed that the forecasting results of the transfer function model is better when compared with other classical forecasting models such as Vector Autoregressive (VAR) and

ARIMA. In addition Thomakos and Geurard [5] who examined unemployment forecasting in St. Louis United States showed that the transfer function model is better when compared to the naive model and ARIMA.

Rainfall data information becomes an important thing that affects the agricultural output to determine the beginning of the growing season. One of the areas in East Java Province that has potential in the agricultural sector is Batu City, so it is important to link rainfall with the agricultural sector in Batu City. The agricultural sector is a leading sector which is expected to synergize with other sectors such as tourism, trade and industry [6]. Some factors that can affect rainfall are air temperature, humidity, wind speed and cloud [7].

Some previous research on rainfall forecasting is done by Faulina [8] about comparison of accuracy of ensemble ARIMA in Batu City and Tresnawati, et. al. [9] examined the Kalman Filter method with SST Nino 3.4 predictors in Purbalingga. Research of rainfall forecasting on preceding still has a weakness that only involves one variable. While research that has been done by Badan Meteorologi, Klimatologi dan Geofisika (BMKG) in rainfall forecasting uses ensemble method (combination of individual method using concept of merging some forecasting model) and BMA (Bayesian Model Averaging) modelling but the forecasting rarely gives the accurate results. Therefore, in this research will discuss rainfall forecasting involving more than one variable.

Air temperature, humidity, wind speed and clouds can be used to model and forecast rainfall using the transfer function model. In this study, rainfall is a response variable (output series), while air temperature, air humidity, wind speed and cloud are predictor variables (input series). The transfer function model used in this research is called the multi-input transfer function model, because the predictor variable or input series is used more than one variable (input). The existence of factors that affect the rainfall as a predictor variable (input series) is expected to provide accurate forecasting results.

FUNDAMENTAL THEORIES

The steps of formation of the transfer function model consists of five steps, namely prewhitening process of input series and output series, identification of impulse response function and transfer function, identification of noise model, prediction and testing of parameter significance of transfer function model, and diagnostic test of transfer function model [10].

Prior to the prewhitening process, the first step is to prepare the input series and the output series. In this preparation step, the identification of time series plots, stationarity test and ARIMA modeling of input series and output series is conducted. Stationarity test consists of two parts, namely stationary to the variance and the mean. Stationarity test against variance can be seen from the Box-Cox plot and the value of λ . The data is said to be stationary to the range if the value of λ is equal to or near 1. The non-stationary time series data can be overcome by doing the Box-Cox transformation. While the stationarity test against the mean can be seen from the autocorrelation plot (ACF). The stationary data on the mean on the ACF plot will not be significant after the second or third lag. Time series data that is not stationary to the mean can be overcome by differencing to become a stationary series. After obtaining a stationary series then the next step is to identify the ARIMA model in the input sequence. The identification of ARIMA model is performed on time series data which has been stationary to the variance and the mean by looking at ACF and PACF plots. The ARIMA model will be used in the prewhitening process of the input series and output series.

The first step of identification of the impulse response function and transfer function is the calculation phase of cross correlation and the autocorrelation of the input series and the prewhitened output series. The cross correlation function (CCF) of the input series x_t and the output series y_t between the k -lags is expressed by $\rho_{xy}(k) = \frac{\gamma_{xy}(k)}{\sigma_x \sigma_y}$ [10]. The second step is the direct prediction of the impulse response weight used to establish the order of the model of the defined transfer function $\hat{v}_k = \frac{C_{\alpha\beta}(k)}{S_{\alpha}^2} = \hat{\rho}_{\alpha\beta}(k) \frac{S_{\beta}}{S_{\alpha}}$ [2]. The last step is the assignment of values (b, r, s) of the transfer function model, b, r and s are three keys of parameter in the transfer function model.

Determining the value of b , r , s can be done by identifying the cross-correlation values of the sample or by directly estimating the weight of the impulse response.

The phase identification model of noise series consists of two steps, the first is the initial estimate of the noise series (n_t) declared $\hat{n}_t = y_t - \frac{\hat{\omega}(B)}{\hat{\delta}(B)} B^b x_t$ [10] and the second is the ARIMA modeling ($p_n, 0, q_n$) for the noise series (n_t).

The estimation of transfer function model parameters used maximum likelihood method and significance test of transfer function model parameters using t test statistic.

The last formation step of the transfer function model is the diagnostic test of the transfer function model which is done by testing the remaining autocorrelation and cross-correlation test (CCF) between the input and the batch input series using the Ljung-Box (Q) test statistic.

Multi-input transfer function model with single output is formulated as follows:

$$y_t = \sum_{j=1}^k v_j(B)x_{jt} + n_t$$

where y_t is output series that has been stationary, x_{jt} is a j-inputs series that have been stationary, n_t is a noise series and $v_j(B)$ is the transfer function for the j-input sequence (x_{jt}).

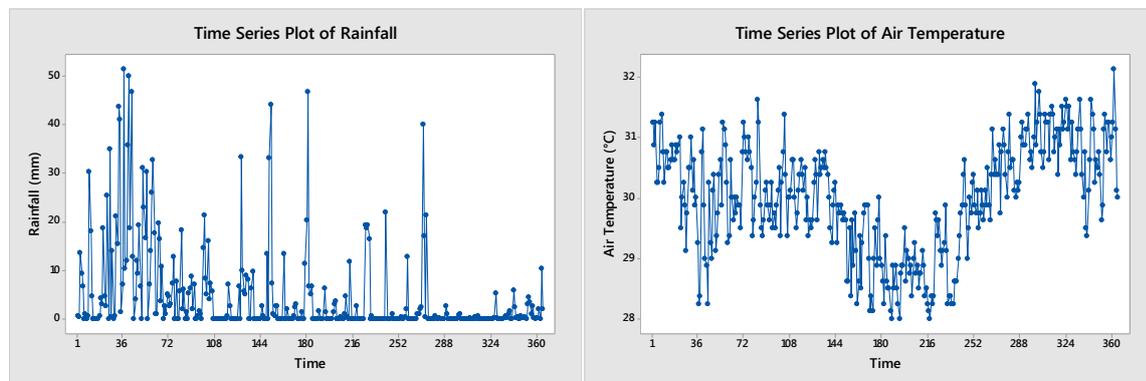
RESULTS AND DISCUSSION

This research used daily secondary data of rainfall (Y_t), air temperature (X_{1t}), humidity (X_{2t}), wind speed (X_{3t}) and cloud (X_{4t}) in Batu City obtained from the web www.worldweatheronline.com. Daily data for the period of January 2016-December 2016 was used as training data to form a multi-input transfer function model. While the daily data for January 2017 period was used as data testing for model validation.

The time series plot of each input series and output series is shown in Figure 1.

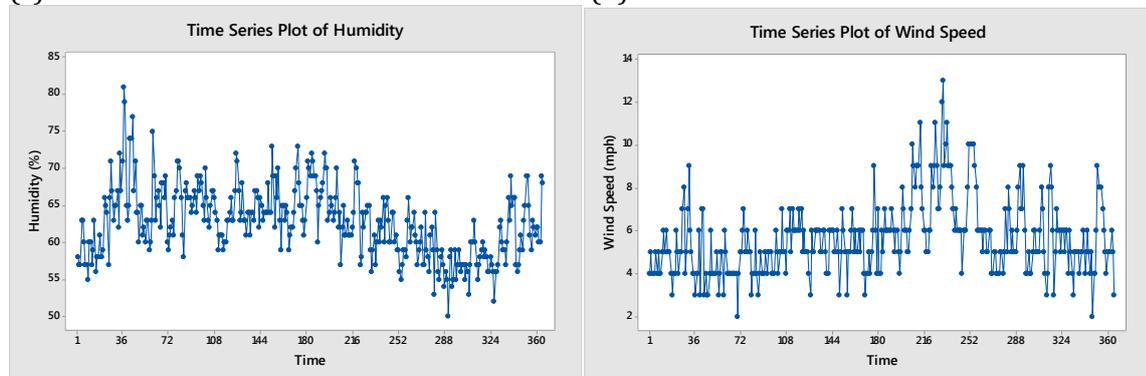
(a)

(b)



(c)

(d)



(e)

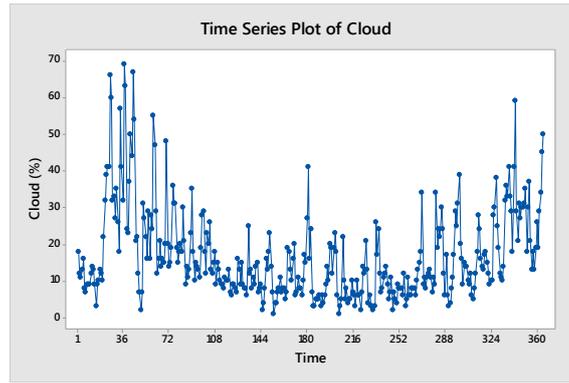


Figure 1. (a) Time series plot of rainfall (b) Time series plot of air temperature (c) Time series plot of humidity (d) Time series plot of wind speed (e) Time series plot of cloud

Based on the time series plot each input series and the output series indicated that the rainfall, air temperature, humidity, wind speed and cloud daily throughout the year 2016 has a pattern up and down and does not show the trend pattern. The stationarity test results to the variance indicated that each input series and output series are not stationary, which is shown by the Box-Cox plot and the λ values of each input and output series are not close to 1. Then, the Box-Cox transformation is done for each the input series and the output series. While the results of stationary test against the mean indicated by ACF and PACF plots also showed that each input series and output series has not been stationary to the mean. Therefore, differencing is done once for each of input and output series so it becomes stationary. Furthermore, the results of ARIMA model identification for the input series obtained ARIMA (1,1,2) for the air temperature input series (X_{1t}), ARIMA ([3,5,7],1,2) for air humidity input series (X_{2t}), ARIMA (1,1,1) for wind speed input series (X_{3t}) and ARIMA (6,1,1) for cloud input series (X_{4t}).

Based on the ARIMA model obtained from each input series, it can be determined prewhitening model on each input series and output series, such as:

$$\alpha_{1t} = x_{1t} - 0.26995x_{1t-1} + 0.63921\alpha_{1t-1} + 0.21176a_{1t-2} \quad (1)$$

$$\alpha_{2t} = x_{2t} + 0.39196x_{2t-1} - 0.24251x_{2t-2} + 0.19483x_{2t-3} + 0.18366x_{2t-5} + 0.36474x_{2t-7} + 0.31646\alpha_{2t-1} + 0.72181\alpha_{2t-2} \quad (2)$$

$$\alpha_{3t} = x_{3t} - 0.51882x_{3t-1} + 0.94334\alpha_{3t-1} \quad (3)$$

$$\alpha_{4t} = x_{4t} - 0.58649x_{4t-1} + 0.14606x_{4t-2} + 0.42116x_{4t-3} + 0.42977x_{4t-4} + 0.16236x_{4t-5} + 0.34483x_{4t-6} + 0.90410\alpha_{4t-1} \quad (4)$$

Equations (1), (2), (3), and (4) are prewhitening models for the input air temperature series, humidity, wind speed and cloud. The prewhitening model for the output series is the same as the input series, only α_{jt} and a_{jt} are replaced by β_{jt} , x_{jt} is replaced by y_{jt} where j is the index of the input series.

The results of the cross-correlation and the autocorrelation of the prewhitened input and output series are shown in the cross-correlation plot of each input air temperature, air humidity, wind speed, cloud series with prewhitened rainfall output are presented in Figure 2.

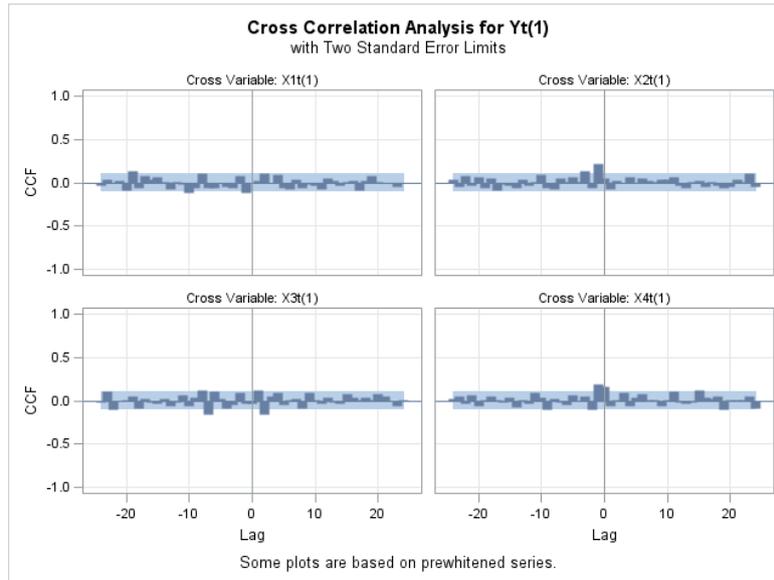


Figure 2. Crosscorrelation Plot of Prewhitened Input Series and Output Series

Based on the results of cross-correlation and the weight of the impulse response for each input series with prewhitened output series, it can be identified the values of b , s , r for each input series. The value of b , s , r for the air temperature input series and the rainfall output series are ($b = 0, s = 0, r = 0$), for the humidity input series and the rainfall output series are ($b = 23, s = 0, r = 0$), while for the wind speed input series and the rainfall output series are ($b = 1, s = 2, r = 0$) and for the cloud input series and the rainfall output series are ($b = 0, s = 0, r = 0$).

The noise series model (n_t) obtained based on the value of the impulse response weight is formulated as the following:

$$n_t = y_t - \hat{y}_t = y_t - \sum_{j=1}^k \hat{v}_j(B)x_{jt}$$

$$\begin{aligned} n_t = y_t - \{ & (-0,1166)x_{1t} + \dots + (-0,8663)x_{1t-23} \} - \{ (0,1391)x_{2t} + \dots + (0,3001)x_{2t-23} \} \\ & - \{ (-0,2078)x_{3t} + \dots + (-0,4434)x_{3t-23} \} \\ & - \{ (0,1681)x_{4t} + \dots + (0,0439)x_{4t-23} \} \end{aligned}$$

Based on the equation we can get the value $n_{24}, n_{25}, \dots, n_{365}$ where the value $n_1 \dots n_{23}$ is set equal to zero. ARIMA model ($p_n, 0, q_n$) for noise series (n_t) is ARIMA ([5,8],0,2) and can be formulated

$$n_t = \frac{(1 + 0.0378B - 0.9109B^2)}{(1 + 0.72949B - 0.23092B^2 - 0.06788B^3 - 0.01765B^4 + 0.1179B^5 + 0.10773B^8)} a_t$$

Identify the values of b , s , r for the multi-input transfer function model based on the incorporation of b , s , r values from the single input transfer function model in each input series to the output series. The possible multi-input transfer function model is $(b_1, s_1, r_1) (b_2, s_2, r_2) (b_3, s_3, r_3) (b_4, s_4, r_4) (p_n, q_n) = (0,0,0) (23,0,0) (1,2,0) (0,0,0) ([5,8],2)$. The prediction and significance test of the parameter of multi-input transfer function model is shown in Table 1.

Table 1. Significance Test Results Parameter Multi Input Transfer Function Model

| Input Series | Orde (B,S,R) | Parameter | Estimation | T | P-Value | Conclusion |
|--------------|------------------|---------------|------------|------|---------|-------------|
| X_{1t} | $b=0,s=0,r=0$ | ω_{01} | 0.1042 | 3.84 | 0.0089 | Significant |
| X_{2t} | $b=23,s=0,r=0$ | ω_{02} | 0.2459 | 2.03 | 0.0423 | Significant |
| X_{3t} | $b=1,s=2,r=0$ | ω_{03} | 0.4318 | 3.53 | 0.0180 | Significant |
| X_{4t} | $b=0,s=0,r=0$ | ω_{04} | 0.1992 | 4.12 | <.0001 | Significant |

Table 1 shows that all parameters in the multi-input transfer function model are significant, this is based on the p-value of each parameter smaller than the value of α .

The diagnostic results of the multi-input transfer function model performed by testing the residual autocorrelation using the Ljung-Box test are shown below:

Table 2. Diagnostic Test Result of Multi Input Transfer Function Model

| Multi Input Transfer Function Model | Lag | Chi-Square | P-Value | Conclusion |
|---|-----|------------|---------|------------|
| $(b_1, s_1, r_1) (b_2, s_2, r_2)$ | 12 | 2.80 | 0.5921 | Fit |
| $(b_3, s_3, r_3) (b_4, s_4, r_4)$ $(p_n, q_n) = (0,0,0)$ | 18 | 7.12 | 0.7142 | |
| $(23,0,0) (1,2,0) (0,0,0)$ $([5,8],2)$ | 24 | 9.66 | 0.8840 | |

Based on the Ljung-Box test, a model is said to be feasible if the whole lag of a model yields a p-value greater than $\alpha = 0.05$. The result of model test using Ljung-Box test stated that the multi input transfer function model with single output is feasible to use.

So the model of multi input transfer function with single output for rainfall forecasting in Batu City stated

$$y_t = 0.10418x_{1t} + 0.24597x_{2t-23} + 0.43178x_{3t-1} + 0.19916x_{4t} + n_t$$

The model shows that rainfall on a certain day is affected by air temperature and cloud on that day, air humidity in the previous 23 days and wind speed in the previous day.

The precision of forecasting was measured by calculating the MAPE (Mean Absolute Percentage Error) statistic. MAPE statistic for testing data was 16.35%. The results of model validation using t paired test showed that the value of forecasting with actual value was not significantly different.

CONCLUSION

The results of analysis showed that based on the multi input with single output transfer function model, the rainfall in Batu City on certain days is affected by air temperature and cloud on that day, humidity in the previous 23 days, and wind speed in the previous day. The results of rainfall forecasting in Batu City using multi input with single output transfer function model is accurate based on the result of model validation using t test and MAPE statistic that less than 20%.

REFERENCES

- [1] G. Box, G. Jenkins dan J. Reisel, *Time Series Analysis Forecasting and Control*, 4th penyunt., New Jersey: Wiley, 2015.
- [2] S. Makridakis dan S. W. a. V. McGee, *Metode dan Aplikasi Peramalan*, 2nd penyunt., Jakarta: Erlangga, 1988.
- [3] C. D. Tankersley, G. W. D dan H. K, "Comparison of Univariate and Transfer Function Models of Groundwater Fluctuations," *Water Resources Research*, no. 29, pp. 3517-3533, 1993.
- [4] P. Edlurd dan K. S, "Forecasting The Swedish Unemployment Rate VAR vs Transfer Function Modelling," *International Journal of Forecasting*, no. 9, pp. 61-76, 1995.
- [5] D. D. Thomakos dan G. J. B, "Naive, ARIMA, Nonparametric, Transfer Function and VAR Models: A Comparison of Forecasting Performance," *International Journal of Forecasting*, no. 20, pp. 53-67, 2004.
- [6] BPS Kota Batu, *Kota Batu dalam Angka*, Batu: BPS Kota Batu, 2015.
- [7] E. Wilson, *Hidrologi Teknik*, 4th penyunt., Jakarta: Erlangga, 1993.
- [8] R. Faulina, "Perbandingan Akurasi Ensemble ARIMA dalam Peramalan Curah Hujan di Kota Batu, Malang, Jawa Timur," *Jurnal Matematika, Sains dan Teknologi*, no. 15, pp. 75-83, 2014.
- [9] R. Tresnawati, T. A. Nuraini dan W. Hanggoro, "Prediksi Curah Hujan Bulanan Menggunakan Metode Kalman Filter dengan Prediktor SST Nino 3.4 Diprediksi," *Jurnal Meteorologi dan Geofisika*, no. 11, pp. 108-119, 2010.
- [10] W. Wei, *Time Series Analysis: Univariate and Multivariate Methods*, 2nd penyunt., Addison-Wesley Publishing Co, 2006.