JAMBURA JOURNAL OF MATHEMATICS
Jambura J. Math. Vol. 4, No. 1, pp. 104–118, January 2022
Journal Homepage: http://ejurnal.ung.ac.id/index.php/jjom
DOI: https://doi.org/10.34312/jjom.v4i1.11975

Assessing Forecasting Performance of Daily Mean Temperature at 1^{st} and 2^{nd} Perak Station Surabaya Using ARIMA and VARIMA Model with Outlier Detection

Taly Purwa[1*], Barbara Ngwarati[2]

[1]*Badan Pusat Statistik Province of Bali, Jl. Raya Puputan No. 1, Denpasar, 80226, Indonesia*
[2]*United College of Education (UCE), Bulawayo, 1156, Zimbabwe*
[*]*Corresponding author. Email: taly@bps.go.id*

ABSTRACT

Air temperature is an important data for several sectors. The demand of fast, exact and accurate forecast on temperature data is getting extremely important since it is useful for planning of several important sectors. In order to forecast mean daily temperature data at 1^{st} and 2^{nd} Perak BMKG Station in Surabaya, this study used the univariate method, ARIMA model and multivariate method, VARIMA model with outlier detection. The best ARIMA model was selected using in-sample criteria, i.e. AIC and BIC. While for VAR model, the minimum information criterion namely AICc value was considered. The RMSE values of several forecasting horizons of out-sample data showed that the overall best model for mean daily temperature at 1^{st} and 2^{nd} Perak Station was the multivariate model, i.e. VARX (10,1) with four outliers incorporated in the model, indicated that it was necessary to consider the temperature from the nearest stations to improve the forecasting performance. This study recommends performing the overall best model only for short term forecasting, i.e. two weeks at maximum. By using the one week-step ahead and one day-step ahead forecasting scheme, the forecasting performance is significantly improved compared to default the k-step ahead forecasting scheme.

Keywords:
Daily Temperature; Forecasting; Multivariate Time Series; Outlier Detection; Vector Autoregression

How to Cite:

T. Purwa and B. Ngwarati, "Assessing Forecasting Performance of Daily Mean Temperature at 1^{st} and 2^{nd} Perak Station Surabaya Using ARIMA and VARIMA Model with Outlier Detection", *Jambura J. Math.*, vol. 4, No. 1, pp. 104–118, 2022, doi: https://doi.org/10.34312/jjom.v4i1.11975

1. Introduction

One important element in weather information is temperature data, i.e. minimum, maximum, and average temperatures. Temperature data in a day is very fluctuating. The difference in temperature conditions in a day is affected by changes in the intensity of the sun's heat that reaches the earth's surface [1]. The recent assessment studies on the impact of climate change indicated that temperature changes affect several sectors

e-ISSN: 2656-1344
Received: 9 November 2021 | Accepted: 6 December 2021 | Online: 1 January 2022

directly i.e. water resources, agriculture, vegetation and tourism [2]. Moreover, the temperature change also plays important role for transportation, industry, and electricity. For example, temperature data is needed by the air transportation sector because the temperature is strongly related to air pressure and wind speed which can affect the flight. The temperature condition also affects the demand and supply of electricity. Temperature conditions that undergo continuous changes are one indicator of climate change that also can affect the environment and health.

Meteorological, Climatological, and Geophysical Agency of Indonesia (BMKG) is required to provide fast, precise and accurate temperature data. In addition, it is also necessary to have temperature forecast data for the next few days that are useful for planning of several important sectors. One of the most commonly used forecasting methods is the Autoregressive Integrated Moving Average (ARIMA) method introduced by Box, *et.al* [3]. This method is known as the univariate method which states that forecasting a variable is only done by relying on the past information of a variable. This forecasting process seems not suitable with actual condition since the temperatures between stations have highly intensity of relationship that can affect each other.

Therefore, in the forecasting process of temperatures on each station, it is necessary to consider the temperature from nearest station called forecasting with a multivariate method. The multivariate methods that are widely used for forecasting are Vector ARIMA (VARIMA). The VARIMA method can explain the effects between several variables simultaneously. By taking into account the effect of other variables, it is expected that the forecast result can be more accurate.

The previous study, Machmudin and Ulama [4] used ARIMA and Artificial Neural Network (ANN) to forecast daily temperature data at 2^{nd} Perak Station, Surabaya. Using the Mean Absolute Percentage Error (MAPE) to measure forecasting performance, the best model to forecast temperature is ANN model. Other study by Ustaoglu, *et.al* [2] also performed ANN algorithm with three different method, i.e. radial basis function (RBF), feed-forward back propagation (FFBP) and generalized regression neural network (GRNN), to forecast daily mean, maximum and minimum temperature of Geyve and Sakarya, Turkey. Study from Liu, *et.al* [5] utilized VAR model to forecast three important weather variables, i.e. temperature, solar radiation and wind speed, for 61 cities around the United States. The results showed that VAR model was suitable for short–term forecasting. The study from Naz [6], performed ARIMA, Exponential Smoothing (ETS), cubic splines and VAR model to forecast maximum daily temperature data at Umea, Swedish. The VAR model did not improve the forecast significantly and ARIMA model was the best model to forecast one-step ahead maximum daily temperature data.

Based on this background, this study seeks to assess the forecasting performance of daily mean temperature data at 1^{st} and 2^{nd} Perak Station, Surabaya using the univariate method, i.e. ARIMA and also multivariate method, i.e. VARIMA, with outlier detection. The outlier detection needs to be performed since the outlier that usually exist in the weather data can greatly affect forecasting accuracy [7]. The forecasting performance for in-sample and several horizons of out-sample data from the two models are compared. The mean-based forecasting also employed as a benchmark to calculate RMSE reduction that produced by each model. From the results of these performance comparison, the overall best model is determined. Finally, using the overall best model, this study forecast the out-sample data using three different schemes, i.e. the default k-step ahead, one week-step ahead and one day-step ahead, to observe the improvement of the forecasting performance.

2. Methods

2.1. Source of Data

This study utilizes two series of secondary data, i.e. the daily mean temperature data at 1^{st} and 2^{nd} Perak Station, Surabaya from January 1^{st}, 2017 to October 31^{st}, 2018 (669 observations) obtained from Indonesian Meteorological, Climatological, and Geophysical Agency (BMKG).

2.2. Analytical Procedures

The analytical procedures used in this study are: first, check for missing values in the data and perform imputation for missing values data. Second, divide the data at 1st and 2nd Perak Station into in-sample data (training data), from January 1st, 2017 to August 31st, 2018 (608 observations), for modeling and out-sample data, from September 1st, 2018 to October 31st, 2018 (61 observations), for forecasting and selecting the overall best model.

Next, check for stationarity in variance and mean condition for each of the in-sample data. The data must meet the stationary in variance and mean condition before modeling process. If the lambda value $\lambda < 1$, then the data is not stationary in variance hence needs to be transformed using Box-Cox's power transformation [8]. Non-stationary in mean indicated by significant plot of Autocorrelation Function (ACF) or Partial Autocorrelation Function (PACF) that decays very slowly hence the differencing process must be applied to the data.

Perform ARIMA method for each in-sample data. The general form of ARIMA with p order Autoregressive (AR) process, q order Moving Average (MA) process and d differencing level or ARIMA (p,d,q) according to [3] is as follows:

$$\phi_p(B)\left(1-B^d\right)Z_t=\theta_q(B)a_t \tag{1}$$

where

$$\phi_p(B)=\left(1-\phi_1 B-\phi_2 B^2-\cdots-\phi_p B^p\right),$$
$$\theta_q(B)=\left(1-\theta_1 B-\theta_2 B^2-\cdots-\theta_q B^q\right),$$

and a_t with $E(a_t)=0$, $Var(a_t)=\sigma_a^2$, and $Cov(a_t,a_{t+k})=0$, $k\neq 0$. The model identification process performed by observing the ACF and PACF plot of stationary data is applied as described by Wei [9]. Select alternative ARIMA models that satisfy three following condition, i.e. all parameters in the models are significant, white noise and normally distributed residuals. Residuals which are not white noise or normally distributed according to the Portmanteau and Kolmogorov-Smirnov test, respectively, will be handled with ARIMAX model. The violation of assumptions usually caused by the presence of outliers in the data. These outliers need to be incorporated in the model to make residuals white noise or normally distributed. There are two type of outliers, i.e. additive outlier (AO) and innovational outlier (IO) [9]. An ARIMA model with AO is define as

$$Z_t = \omega I_t^{(T)} + \frac{\theta_q(B)}{\phi_p(B)} a_t \tag{2}$$

where $I_t^{(T)} = 1, t = T$ and $I_t^{(T)} = 0, t \neq T$. While, an ARIMA model with IO is define as

$$Z_t = \frac{\theta_q(B)}{\phi_p(B)} \left(\omega I_t^{(T)} + a_t \right) \tag{3}$$

where $I_t^{(T)} = 1, t \geq T$ and $I_t^{(T)} = 0, t < T$. Generally, the ARIMA model with k several outliers is define as

$$Z_t = \sum_{j=1}^{k} \omega_j v_j(B) I_t^{(T)} + \frac{\theta_q(B)}{\phi_p(B)} a_t \tag{4}$$

where $v_j(B) = 1$ for an AO and $v_j(B) = \frac{\theta_q(B)}{\phi_p(B)}$ for an IO at time $t = T_j$. The best ARIMA or ARIMAX model is chosen using in-sample criteria, i.e. AIC and BIC.

Next, perform VARIMA method for both in-sample data. This multivariate time series model has an advantage to explain the relationships between multiple sets of time series data. According to Wei [9] and Tsay [10], the general model of VARIMA(p, d, q) is as follows:

$$\boldsymbol{\Phi}_p(B) \boldsymbol{D}(B) \boldsymbol{Z}_t = \boldsymbol{\theta}_q(B) \boldsymbol{a}_t \tag{5}$$

where $\boldsymbol{Z}_t = \begin{bmatrix} Z_{1,t} & Z_{2,t} & \cdots & Z_{m,t} \end{bmatrix}'$, $\boldsymbol{a}_t = \begin{bmatrix} a_{1,t} & a_{2,t} & \cdots & a_{m,t} \end{bmatrix}'$, diagonal matrix of differencing operator $\boldsymbol{D}(B)$ with diagonal values $(1-B)^{d_1}$, $(1-B)^{d_2}$, ..., $(1-B)^{d_m}$, $\boldsymbol{\Phi}_p(B) = 1 - \boldsymbol{\Phi}_1 B - \boldsymbol{\Phi}_2 B^2 - \cdots - \boldsymbol{\Phi}_p B^p$, and $\boldsymbol{\theta}_q(B) = 1 - \boldsymbol{\theta}_1 B - \boldsymbol{\theta}_2 B^2 - \cdots - \boldsymbol{\theta}_q B^q$ with

$$\boldsymbol{\Phi}_p = \begin{bmatrix} \phi_{p\_11} & \phi_{p\_12} & \cdots & \phi_{p\_1m} \\ \phi_{p\_21} & \phi_{p\_22} & \cdots & \phi_{p\_2m} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{p\_m1} & \phi_{p\_m2} & \cdots & \phi_{p\_mm} \end{bmatrix}$$

and

$$\boldsymbol{\theta}_q = \begin{bmatrix} \theta_{q\_11} & \theta_{q\_12} & \cdots & \theta_{q\_1m} \\ \theta_{q\_21} & \theta_{q\_22} & \cdots & \theta_{q\_2m} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{q\_m1} & \theta_{q\_m2} & \cdots & \theta_{q\_mm} \end{bmatrix}$$

The next process is model identification or determine the number of lagged values based on Akaike Information Criterion (AIC) produced by VARIMA model as discussed by Lutkepohl [11]. In the VARIMA model estimation process, non-significant estimations of the parameters can be restricted (assumed as constant value of 0). Then proceed the diagnostic checking to ensure that the residuals a_t are white noise according to the corrected AIC (AICc) value and satisfy multivariate normal distribution assumptions using squared generalized distance [12], i.e. $d_t^2 = (\boldsymbol{a}_t - \overline{\boldsymbol{a}})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{a}_t - \overline{\boldsymbol{a}})$, with $\boldsymbol{a}_t$ is $n \times m$ matrix of residuals and $\boldsymbol{\Sigma}$ is variance covariance matrix of residuals. The m-variate normality is indicated if the number of $d_t^2 \leq \chi^2_{(0.5,m)}$ is at least 50% and the chi-square plot of ordered d_t^2 vs. $\chi^2_{((n-t+0.5)/n,m)}$ is reasonably straight compared to diagonal line (slope 1). If the residuals are not white noise or follow multivariate normal distribution then perform outlier detection for multivariate data using the T2 Hotelling statistics [13], that has same formula with the d_t^2, as follows

$$T_i^2 = (\boldsymbol{a}_t - \overline{\boldsymbol{a}})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{a}_t - \overline{\boldsymbol{a}}) \tag{6}$$

If T_i^2 is below the lower control limit (LCL), 0 or greater than upper control limit (UCL), $\chi^2_{(\alpha,m)}$, with m is number of time series data, then the i-th observation are categorized as outliers. In the VARIMAX model, these outliers are incorporated as exogenous variable, i.e. as dummy variables, with the general model of VARIMAX (p, q, s) as follow

$$\boldsymbol{\Phi}_p(B)\,\boldsymbol{D}(B)\,\boldsymbol{Z}_t = \sum_{j=0}^{s} \boldsymbol{\beta}_j \boldsymbol{X}_{t-j} + \boldsymbol{\theta}_q(B)\,\boldsymbol{a}_t \tag{7}$$

where $\boldsymbol{\beta}_j$ is $m \times r$ matrix of dummy variable's parameter and $\boldsymbol{X}_{t-j}$ is $r \times 1$ vector of dummy variable, with r is number of outliers.

Using the best ARIMA and VARIMA models for both in-sample data from the previous procedures, forecasting for several horizons ahead up until two months, i.e. September and October, would be performed then the overall best model for daily mean temperature data at 1^{st} and 2^{nd} Perak Station are selected based on RMSE. We also calculate the reduction of RMSE for each data by using ARIMA and VARIMA method compared to mean-based forecasting as a benchmark. The final procedure is the forecasting performance inspection of the overall best model from different schemes, i.e. k-step ahead, one week-step ahead and one day-step ahead forecasting.

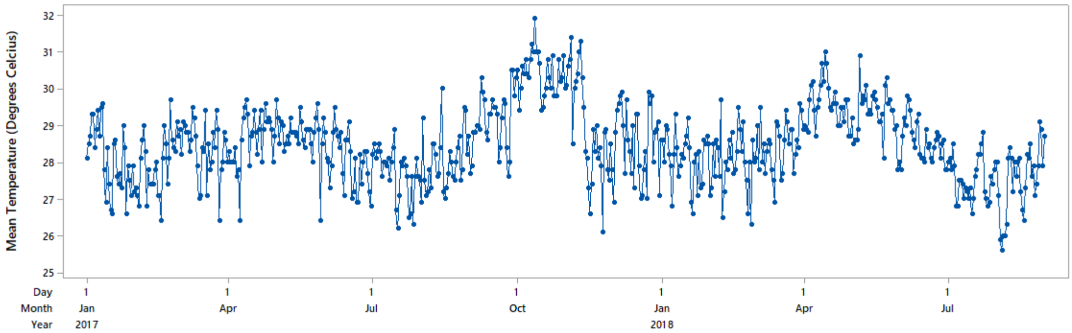
3. Results and Discussion

3.1. Imputation for Missing Values

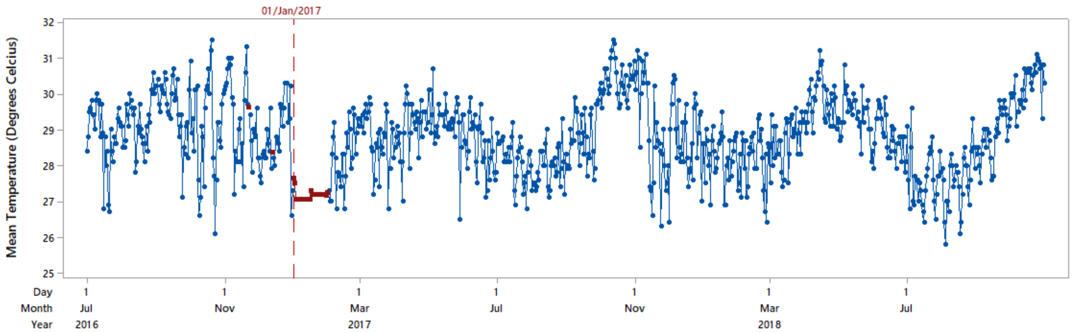
There were no missing values in the daily mean temperature data at 1^{st} Perak Station, however there were 31 missing values (January 1^{st} until 31^{st}, 2017) in the data at 2^{nd} Perak Station. Since these missing values are appear in the beginning of the in-sample data so we involved a previous year (2016) data for imputation process using simple Moving Average [14]. The time series plot for daily mean temperature data at 1^{st} and 2^{nd} Perak Station are presented in Figure 1 (a) and (b), respectively.

3.2. ARIMAX Model

The condition of stationarity in variance for each data is checked by value of λ from Box Cox. The rounded value of λ for mean temperature data at 1^{st} and 2^{nd} Perak Station were 4 and 3, respectively. The 95% confidence level of λ for mean temperature data at 1^{st} Perak Station did not contain 1, indicated non-stationary condition. But transformation



(a)



(b)

Note: *red line shows imputed values*

Figure 1. Time series plot of daily mean temperature data at 1^{st} (a) and 2^{nd} (b) Perak station

was not performed since it would increase the variance of the data. For the 2^{nd} Perak Station, the 95% confidence level of λ contain 1 then the data already meet stationary in variance condition.

The ACF and PACF plot of daily mean temperature at 1^{st} and 2^{nd} Perak Station (Figure 2 (a) and (b)) show that the data do not meet stationarity in mean condition hence we need to perform regular differencing ($d=1$) to the data.

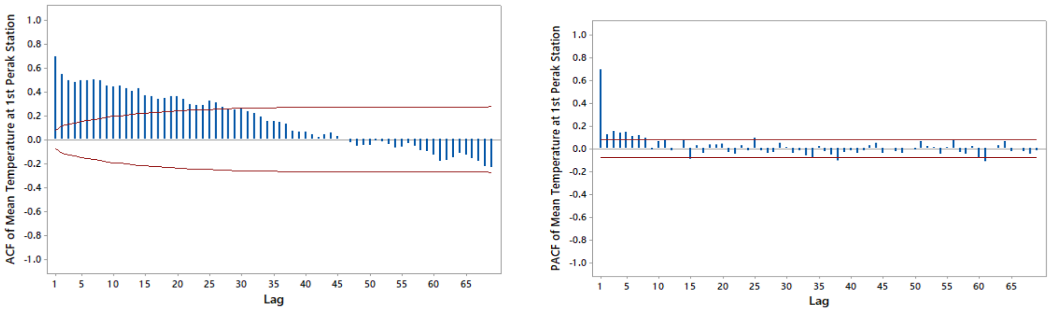
After regular differencing, the ACF and PACF plot presented in Figure 3 (a) and (b) have indicated stationarity in mean condition. According to these ACF and PACF plot, we can identify the ARIMA model and estimate the parameters.

For daily mean temperature data at 1^{st} Perak Station, there are three ARIMA models that have all significant parameters in the models with $\alpha = 0,05$, white noise until lags 48 and normally distributed residuals based on Kolmogorov-Smirnov test as presented in Table 1.

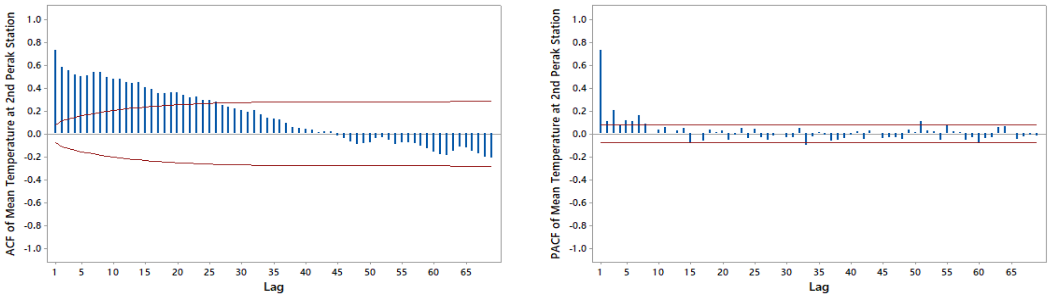
The best ARIMA model for daily mean temperature data at 1^{st} Perak station was model that has minimum AIC and BIC, i.e. ARIMA (0,1,2) with the following equation,

$$(1-B)\,\dot{Z}_{1,t} = (1-0.5006B-0.2863B^2)a_{1,t} \tag{8}$$

or

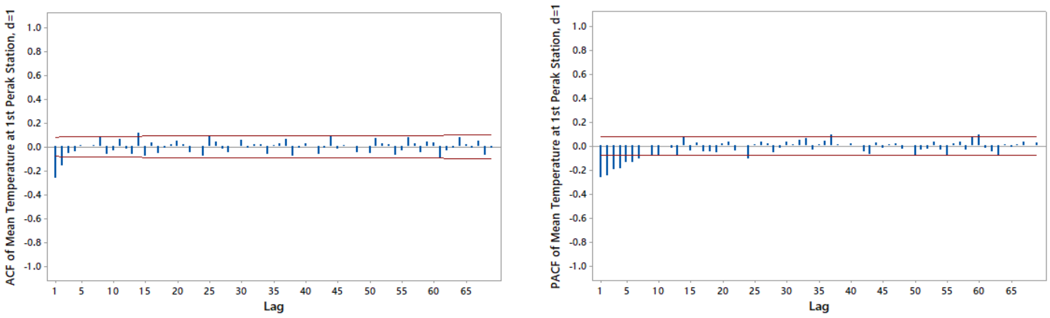


(a)

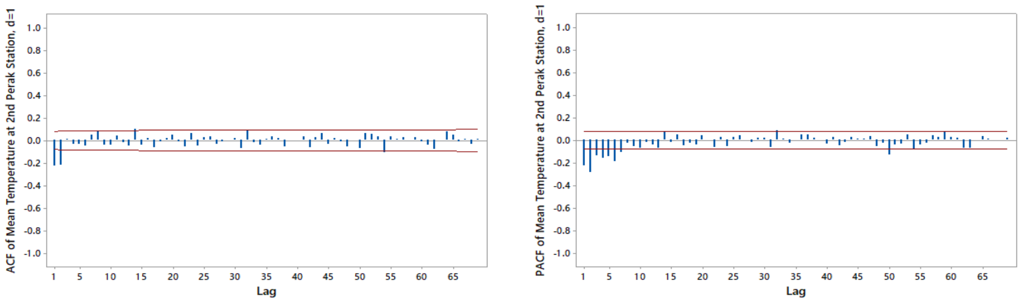


(b)

Figure 2. ACF and PACF plot of daily mean temperature data at 1^{st} (a) and 2^{nd} (b) Perak station



(a)



(b)

Figure 3. ACF and PACF plot of daily mean temperature data at 1^{st} (a) and 2^{nd} (b) Perak station, after regular differencing (d= 1)

Table 1. AIC and BIC values of ARIMA models for daily mean temperature at 1^{st} Perak station

Model	AIC	BIC
ARIMA (0,1,2)	1332.851*	1341.668*
ARIMA (2,1,[2])	1335.107	1348.333
ARIMA (1,1,1)	1333.234	1342.051

Note: *) *The Best Model*

$$Z_{1,t} = Z_{1,t-1} + a_{1,t} - 0.5006a_{1,t-1} - 0.2863a_{1,t-2} \tag{9}$$

For daily mean temperature data at 2^{nd} Perak Station, there were 5 ARIMA models that have all of the significant parameters and white noise residuals however all of these residuals were not normally distributed. Then modeling using ARIMAX model is performed. There are five alternative models obtained in Table 2.

Table 2. AIC and BIC values of ARIMAX models for daily mean temperature at 2^{nd} Perak station

Model	Outlier	AIC	BIC
ARIMA (7,1,0)	AO = 149, 422, 322, 309, 85, 426, 77, 98, 328, 551 IO = 336, 317, 278, 364, 357	1111.119	1208.107
ARIMA (0,1,2)	AO = 149, 322, 422, 85, 309, 364, 328 IO = 317, 336, 270	1142.766*	1195.668
ARIMA (2,1,1)	AO = 149, 322, 422, 364, 85, 309, 328 IO = 317, 336, 270	1144.279	1201.59
ARIMA (2,1,[2])	AO = 149, 322, 364, 309, 85, 328, 98 IO = 317, 336, 270	1152.422	1209.733
ARIMA ([1,2,5],1,[2])	AO = 149, 322, 422, 85, 309, 328 IO = 317, 336, 270, 552	1148.765	1210.484

Note: *) *The Best Model*

The best ARIMAX model for daily mean temperature data at 2^{nd} Perak station is ARIMA (0,1,2) with 10 outliers with the following equation,

$$\begin{aligned} Z_{2,t} = {} & -2,2983I_{2,t}^{149} + 2,1403I_{2,t}^{322} - 2,1226I_{2,t}^{422} - 1,8916I_{2,t}^{85} - 1,8149I_{2,t}^{309} - 1,8204I_{2,t}^{364} \\ & -1,7592I_{2,t}^{328} + \frac{(1 - 0,4415B - 0,3634\mathrm{B}^2)}{(1-B)}(-2,4066I_{2,t}^{(317)} + 1,6972I_{2,t}^{336} + 1,5944I_{2,t}^{270} + a_{2,t}) \end{aligned} \tag{10}$$

3.3. VARIMAX Model

In the model identification process, the VARIMA models are conducted for $p = 0, 1, 2, \ldots, 10$ and $q = 0, 1, 2, \ldots, 5$ then the AICc values are compared. The model with minimum AICc value, i.e. -2.6489, is obtained by VARIMA model with $p = 9$ and $q = 0$ or VAR (9) model. Next, VAR (9) model is estimated and non-significant parameters are restricted one by one until all significant parameters remain in the model.

According to the d_i^2 value, the residuals follow multivariate normal distribution. Also white noise checking of residuals is performed by modeling the vector of residuals from restricted VAR (9) model and observing the minimum AICc value. When the minimum AICc values is not produced from model with $p = q = 0$, the residuals are not white noise. This condition is caused by the presence of outliers observed from the outliers detection using $T2$ Hotelling statistics that there are 23 outliers with the T_i^2 values more than the UCL, $\chi^2_{(0.99,2)} = 9.2103$ as presented in Figure 4.

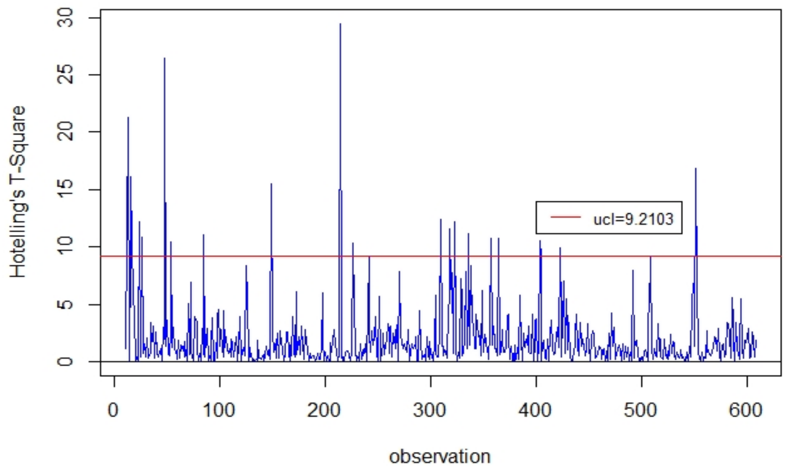


Figure 4. Plot of $T2$ Hotelling statistics of the residuals

In the first iteration, four biggest outliers, i.e. 214^{th}, 48^{th}, 13^{rd}, and 551^{st} observation, are incorporated in the model. The VARX (10,1) model has the minimum AICc in the model identification process. After estimating the model, the restriction for the non-significant parameters are performed. From the modeling of the residuals of restricted model, the minimum AICc value is -2.9099 produced from model with $p = q = 0$ indicating that the residuals are white noise. These residuals also follow multivariate normal distribution since 57.45% of its d_i^2 value are less than or equal to $\chi^2_{(0.5,2)} = 1.3863$ and also have reasonable straight pattern in Figure 5.

The final estimation model is VARX (10,1) model with 4 outliers as follow:

$$\left(1 - \boldsymbol{\Phi}_1 B - \boldsymbol{\Phi}_2 B^2 - \cdots - \boldsymbol{\Phi}_{10} B^{10}\right)(1 - B)\,\boldsymbol{Z}_t = \boldsymbol{\beta}_0 \boldsymbol{X}_t + \boldsymbol{\beta}_1 \boldsymbol{X}_{t-1} + \boldsymbol{a}_t \tag{11}$$

or

$$\begin{aligned} \boldsymbol{Z}_t &= (\boldsymbol{I} + \boldsymbol{\Phi}_1)\boldsymbol{Z}_{t-1} + (\boldsymbol{\Phi}_2 - \boldsymbol{\Phi}_1)\boldsymbol{Z}_{t-2} + (\boldsymbol{\Phi}_3 - \boldsymbol{\Phi}_2)\boldsymbol{Z}_{t-3} + (\boldsymbol{\Phi}_4 - \boldsymbol{\Phi}_3)\boldsymbol{Z}_{t-4} \\ &+ (\boldsymbol{\Phi}_5 - \boldsymbol{\Phi}_4)\boldsymbol{Z}_{t-5} + (\boldsymbol{\Phi}_6 - \boldsymbol{\Phi}_5)\boldsymbol{Z}_{t-6} + (\boldsymbol{\Phi}_7 - \boldsymbol{\Phi}_6)\boldsymbol{Z}_{t-7} + (\boldsymbol{\Phi}_8 - \boldsymbol{\Phi}_7)\boldsymbol{Z}_{t-8} \\ &+ (\boldsymbol{\Phi}_9 - \boldsymbol{\Phi}_8)\boldsymbol{Z}_{t-9} + (\boldsymbol{\Phi}_{10} - \boldsymbol{\Phi}_9)\boldsymbol{Z}_{t-10} - \boldsymbol{\Phi}_{10}\boldsymbol{Z}_{t-11} + \boldsymbol{\beta}_0 \boldsymbol{X}_t + \boldsymbol{\beta}_1 \boldsymbol{X}_{t-1} + \boldsymbol{a}_t \end{aligned} \tag{12}$$

or

$$\begin{bmatrix} Z_{1,t} \\ Z_{2,t} \end{bmatrix} = \begin{bmatrix} 0.2565 & 0.2916 \\ 0 & 0.5632 \end{bmatrix} \begin{bmatrix} Z_{1,t-1} \\ Z_{2,t-1} \end{bmatrix} + \begin{bmatrix} 0.0598 & -0.0753 \\ 0 & -0.0506 \end{bmatrix} \begin{bmatrix} Z_{1,t-2} \\ Z_{2,t-2} \end{bmatrix}$$

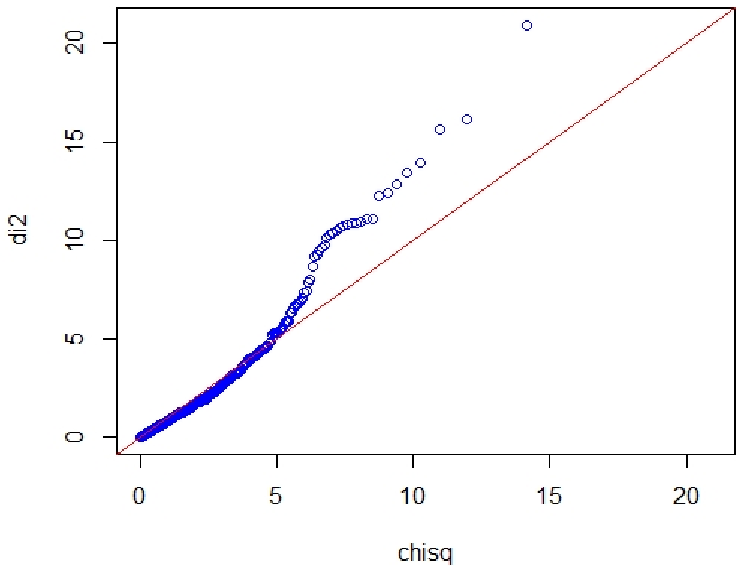


Figure 5. Chi-square plot of residuals squared generalized distance

$$
\begin{aligned}
&+\begin{bmatrix} 0.0999 & -0.0385 \\ -0.1158 & 0.2217 \end{bmatrix}\begin{bmatrix} Z_{1,t-3} \\ Z_{2,t-3} \end{bmatrix}+\begin{bmatrix} 0.1609 & -0.1778 \\ 0.1158 & -0.1384 \end{bmatrix}\begin{bmatrix} Z_{1,t-4} \\ Z_{2,t-4} \end{bmatrix} \\
&+\begin{bmatrix} 0.1105 & 0 \\ 0 & 0.0937 \end{bmatrix}\begin{bmatrix} Z_{1,t-5} \\ Z_{2,t-5} \end{bmatrix}+\begin{bmatrix} 0.1071 & -0.0996 \\ 0 & -0.0063 \end{bmatrix}\begin{bmatrix} Z_{1,t-6} \\ Z_{2,t-6} \end{bmatrix} \\
&+\begin{bmatrix} 0.0210 & 0.0996 \\ 0 & 0.1309 \end{bmatrix}\begin{bmatrix} Z_{1,t-7} \\ Z_{2,t-7} \end{bmatrix}+\begin{bmatrix} 0.0689 & 0 \\ 0 & 0.0729 \end{bmatrix}\begin{bmatrix} Z_{1,t-8} \\ Z_{2,t-8} \end{bmatrix} \\
&+\begin{bmatrix} -0.0832 & 0.0831 \\ -0.1112 & 0.1129 \end{bmatrix}\begin{bmatrix} Z_{1,t-9} \\ Z_{2,t-9} \end{bmatrix}+\begin{bmatrix} 0.0919 & -0.0831 \\ 0.0236 & 0 \end{bmatrix}\begin{bmatrix} Z_{1,t-10} \\ Z_{2,t-10} \end{bmatrix} \\
&+\begin{bmatrix} 0.1067 & 0 \\ 0.0876 & 0 \end{bmatrix}\begin{bmatrix} Z_{1,t-11} \\ Z_{2,t-11} \end{bmatrix}+\begin{bmatrix} 2.2096 & -2.0906 & -1.9226 & 0 \\ 0 & 0 & 0 & 1.6352 \end{bmatrix}\begin{bmatrix} X_{1,t} \\ X_{2,t} \\ X_{3,t} \\ X_{4,t} \end{bmatrix} \\
&+\begin{bmatrix} 0 & 0 & 0 & -1.8328 \\ 0 & 0 & 0 & -1.9330 \end{bmatrix}\begin{bmatrix} X_{1,t-1} \\ X_{2,t-1} \\ X_{3,t-1} \\ X_{4,t-1} \end{bmatrix}+\begin{bmatrix} a_{1,t} \\ a_{2,t} \end{bmatrix} \qquad (13)
\end{aligned}
$$

3.4. Selection of the Overall Best Model

The forecasting performance was measured using RMSE for in-sample and out-sample data (Table 3). For in-sample data, the best model for daily mean temperature at 1^{st} Perak Station is VARX (10,1) and at 2^{nd} Perak Station the best model was ARIMAX. The different results were shown for out-sample data, where the VARX (10,1) model was the best model both for daily mean temperature at 1^{st} and 2^{nd} Perak Station since it has a minimum RMSE for across all horizons. The RMSE values for out-sample data of daily mean temperature at 1^{st} and 2^{nd} Perak Station for both model increases as the forecasting period increases this means that the forecasting results are getting inaccurate for longer horizon. According to the RMSE value across all horizons for out-sample data, the overall best model for daily mean temperature at 1^{st} and 2^{nd} Perak Station is VARX (10,1).

Table 3. RMSE of ARIMA and VAR models for in-sample and out-sample data

Output	Model	In-sample	Out-sample by Horizons			
			One Week	Two Weeks	One Month	Two Months
1^{st} Perak	ARIMA	0.7230	0.6573	0.6809	0.9530	1.6782
	VARX	0.7014*	0.6270*	0.6650*	0.8867*	1.5891*
2^{nd} Perak	ARIMAX	0.6081*	0.5706	0.5817	0.9802	1.7515
	VARX	0.6751	0.5117*	0.5376*	0.8806*	1.6316*

Note: *) *The Best Model*

The forecasting performance from two models for one week and two weeks forecasting horizons outperforms the performance for in-sample data. This condition indicates that the models are suitable for short-term forecasting (two weeks at maximum) as confirmed by the percentage of RMSE that can be reduced by the ARIMA or ARIMAX and VARX model from the error measures generated from mean-based forecasting as a benchmark in Figure 6. This procedure also was conducted in previous study [15]. The RMSE reduction decreases as the forecasting horizon increases for all methods, even significantly decrease for one month forecasting and getting worse for two months forecasting horizon since it has negative values.

3.5. Improving the Forecasting Performance of The Overall Best Model

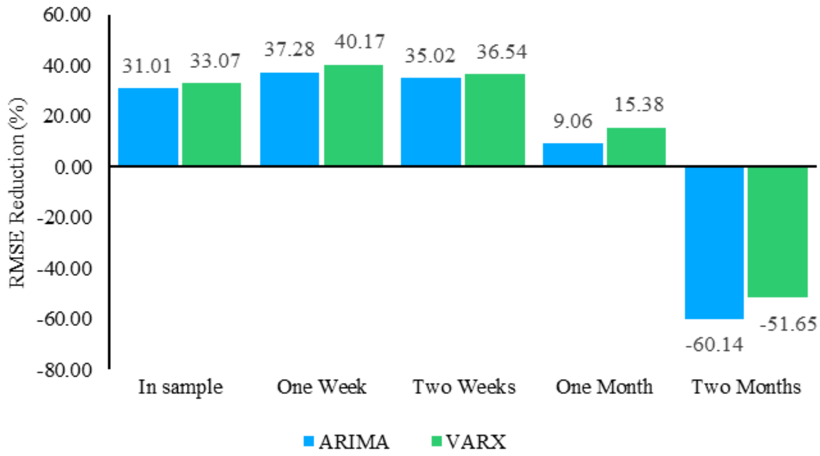
The forecasting scheme for out-sample data used in previous results is k-step ahead. This study also elaborates other schemes, i.e. one week-step ahead and one day-step ahead forecasting, to obtain improvement of forecasting performance from the overall best model. Table 4 shows that generally the k-step ahead produces greater RMSE or has lower accuracy compared to other schemes, especially for one month and two months horizons, since this default k-step ahead accumulates the error terms [16]. The two alternative schemes produce an improvement of forecasting performance, except the one week-step ahead scheme for one week and two weeks horizons that has RMSE equal or greater than k-step ahead scheme. The one day-step ahead scheme can produce the best performance than other schemes for all forecasting horizons since have smallest and relatively constant RMSE value compared to other forecasting schemes.

Table 4. RMSE of the overall best model for out-sample data using several schemes

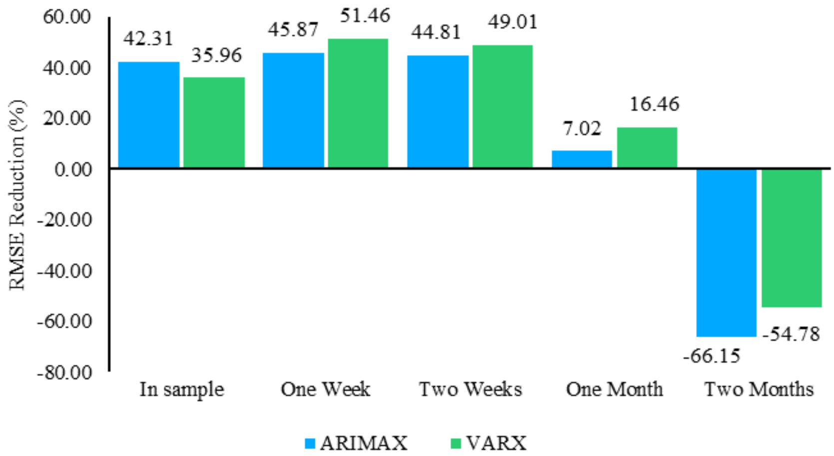
Output	Scheme	Out sample by Horizons			
		OneWeek	Two Weeks	One Month	Two Months
1^{st} Perak	k-step ahead	0.6270	0.6650	0.8867	1.5891
	One week-step ahead	0.6270	0.7160	0.7724	0.6928
	One day-step ahead	0.5132*	0.5754*	0.5709*	0.5858*
2^{nd} Perak	k-step ahead	0.5117	0.5376	0.8806	1.6316
	One week-step ahead	0.5117	0.5729	0.7096	0.6510
	One day-step ahead	0.4207*	0.4493*	0.5000*	0.5316*

Note: *) *The Smallest RMSE*

Figure 7 illustrates the forecasting performance of each scheme and shows that forecasting value from one day-step ahead scheme has more similar pattern with actual



(a)



(b)

Figure 6. RMSE reduction of ARIMA or ARIMAX and VARX models for daily mean temperature at 1^{st} (a) and 2^{nd} (b) Perak station

out-sample data compared to other schemes makes the best forecasting performance as described before. This figure also shows when the models need to be updated as shown by the red vertical dotted line, when the actual value is outside the 95% forecast interval. The overall best model needs to be updated from October 5^{th}, 2018 when using k-step ahead scheme. By using the one day-step ahead scheme, the time for updating is longer, i.e. from October 29^{th}, 2018. While, by using the one week-step ahead scheme, the overall best model does not need to be updated for two months interval of out-sample data.

4. Conclusion

This study uses the ARIMA and VARIMA model, each with outlier detection, to assess the forecasting performance of daily mean temperature data at 1^{st} and 2^{nd} Perak Station, Surabaya. Both the series data consist of trend component that needs to be regularly differenced. The outliers are incorporated as dummy variables in the ARIMA model of the 1^{st} Perak Station data and VARIMA model since its residuals were not normally distributed and not white noise, respectively. According to the RMSE values of several

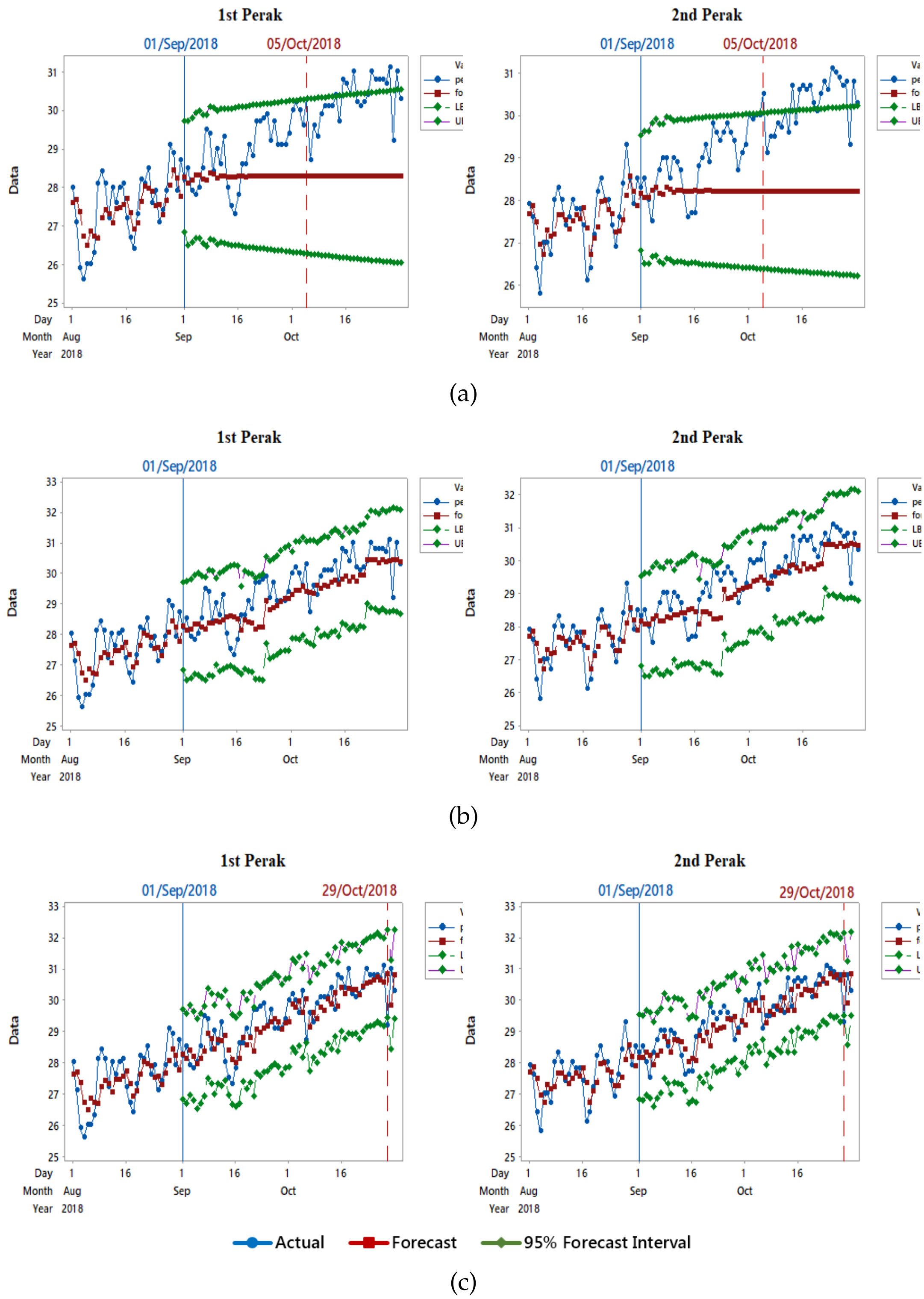


Figure 7. Actual data, forecasting value and 95% forecast interval of k-step ahead (a), one week-step ahead (b) and one day-step ahead (c) forecasting schemes

forecasting horizons of out-sample data, the overall best model for daily mean temperature at 1st and 2nd Perak Station is the multivariate model, i.e. VARX (10,1) model that forecasting results got inaccurate for longer horizon. This condition indicated that by utilizing other data to the model, especially from the nearest location, can improve the forecasting performance for each data.

Using the default k-step ahead forecasting scheme, this study recommends performing the overall best model only for short term forecasting (two weeks maximum) as the RMSE after two weeks are bigger than the RMSE of in-sample data which is also confirmed by significantly decreasing of the RMSE reduction when comparing to the mean-based forecasting as benchmark. Using the same scheme, the model also needs to be updated when forecasting from October 5th, 2018 and afterward. The forecasting performance is significantly improved by using the one week-step ahead and one day-step ahead forecasting schemes with the best performance produced by the one day-step ahead scheme that has relatively constant RMSE and has more similar pattern with actual out-sample data but the model needs to be updated from October 29th, 2018. While the model with the one week-step ahead scheme does not need to be updated for two months interval of out-sample data.

References

[1] BPS (Badan Pusat Statistik), "Statistik Lingkungan Hidup Indonesia 2017," Badan Pusat Statistik, Tech. Rep., 2017.

[2] B. Ustaoglu, H. K. Cigizoglu, and M. Karaca, "Forecast of daily mean, maximum and minimum temperature time series by three artificial neural network methods," *Meteorological Applications*, vol. 15, no. 4, pp. 431–445, dec 2008, doi: http://dx.doi.org/10.1002/met.83.

[3] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis : Forecasting and Control*, 3rd ed. Englewood Cliffs, NJ: Prentice Hall, 1994.

[4] A. Machmudin and B. S. S. Ulama, "Peramalan Temperatur Udara di Kota Surabaya dengan Menggunakan ARIMA dan Artificial Neural Network," *J. SAINS DAN SENI ITS*, vol. 1, no. 1, pp. D–118–D123, 2012.

[5] Y. LIU, M. C. ROBERTS, and R. SIOSHANSI, "A vector autoregression weather model for electricity supply and demand modeling," *Journal of Modern Power Systems and Clean Energy*, vol. 6, no. 4, pp. 763–776, jul 2018, doi: http://dx.doi.org/10.1007/s40565-017-0365-1.

[6] S. Naz, "Forecasting Daily Maximum Temperature of Umeå," Umeå University, Umeå University, Tech. Rep., 2015.

[7] C. Bergmeir and J. M. Benítez, "On the use of cross-validation for time series predictor evaluation," *Information Sciences*, vol. 191, pp. 192–213, may 2012, doi: http://dx.doi.org/10.1016/j.ins.2011.12.028.

[8] G. E. P. Box and D. R. Cox, "An Analysis of Transformations," *Journal of the Royal Statistical Society*, vol. 26, no. 2, pp. 211–252, 1964.

[9] W. W. S. Wei, *Time Series Analysis Univariate and Multivariate Methods*, 2nd ed., New York, 2006.

[10] R. S. Tsay, *Multivariate Time Series Analysis - With R and Financial Applications*. New Jersey: John Wiley and Sons, Inc., 2014.

[11] H. Lutkepohl, *Introduction to Multiple Time Series Analysis*, 2nd ed. Berlin: Springer-Verlag, 1993.

[12] R. . Johnson and D. Winchern, *Applied Multivariate Statistical Analysis*, 6th ed. Upper Saddle River: Pearson Education Inc., 2007.

[13] R. Ödük, "Control Charts for Serially Dependent Multivariate Data," DTU Informatics, Technical University of Denmark, DTU Informatics, Tech. Rep., 2012.

[14] S. Moritz and T. Bartz-Beielstein, "imputeTS: Time Series Missing Value Imputation in R," *The R Journal*, vol. 9, no. 1, pp. 207–218, 2017, doi: http://dx.doi.org/10.32614/RJ-2017-009.

[15] T. Purwa, U. Nafngiyana, and S. Suhartono, "Comparison of ARIMA, Transfer Function and VAR Models for Forecasting CPI, Stock Prices, and Indonesian Exchange Rate: Accuracy vs. Explainability," *MEDIA STATISTIKA*, vol. 13, no. 1, pp. 1–12, jun 2020, doi: http://dx.doi.org/10.14710/medstat.13.1.1-12.

[16] N. Mohamed, M. H. Ahmad, Z. Ismail, and S. Suhartono, "Short Term Load Forecasting Using Double Seasonal ARIMA Model," in *Proceedings of the Regional Conference on Statistical Sciences 2010 (RCSS'10)*, 2010, pp. 57–73.