Rainfall Forecasting Model Using ARIMA and Kalman Filter in Makassar, Indonesia

Sukarna¹*, Elma Yulia Putri Ananda¹, and Maya Sari Wahyuni¹
¹Mathematics Department, Universitas Negeri Makassar, Indonesia

*Email: sukarna@unm.ac.id

Abstract. Many forecasting methods have been used for forecasting rainfall data. Kalman Filter is one of the forecasting methods that could give better forecasts. To our knowledge, the Kalman Filter method has not been used to forecast rainfall data in Makassar, Indonesia. This study aims to provide more precise forecasts for rainfall data in Makassar, Indonesia by using Autoregressive Integrated Moving Average (ARIMA) and Kalman Filter methods. Rainfall data from January 2010 to December 2020 were used. The best model selection is based on the smallest Mean Absolute Percentage Error (MAPE) value. The results showed that the best ARIMA model is ARIMA(0,1,1)(0,1,1)¹² with MAPE is 111.48, while MAPE value by using the Kalman Filter algorithm is 47.00 indicating that Kalman Filter has better prediction than ARIMA model.

1. Introduction
Geographically, Makassar City is located at 119°24′17″ East Longitude and 5°8′6″ South Latitude [1]. Makassar has a tropical climate with an average temperature ranging from 26.2°C – 29.3°C with humidity of 77 percent and an average wind speed of 5.2 knots [2,3]. In general, the rainfall in Makassar is quite varied throughout the year. Makassar City experiences the rainy season from November to April, while the dry season experiences from May to October. When the rainfall is quite high, several areas in Makassar City have puddles of water so that some areas experience flooding. Meanwhile, when the rainfall is low, some areas experience drought caused by low soil absorption capacity [4].

Several studies on rainfall data have been conducted in Indonesia using different methods. One study used the Markov Chain model to forecast the rainfall data in Makassar [5]. Another study used Vector Autoregressive Neural Network to forecast rainfall data in East Surabaya [6]. The comparison of Artificial Neural Network (ANN) methods to forecast rainfall data has also been evaluated [7]. Forecasting rainfall data in Malang Regency using VAR-NN and GSTAR-NN [8], Comparison of Rainfall Forecasting using Bayesian Model Averaging and Kalman Filter Method [9] and application of Kalman Filter model in predicting rainfall in Kubu Raya Regency [10] have been also studied. Although several studies of rainfall forecasting have been done, to our knowledge, the rainfall forecasting model using ARIMA and Kalman Filter in Makassar, Indonesia has not been explored yet.

The rainfall in Makassar City which varies every year has an impact on human life such as agricultural production, fisheries, plantations, and aviation [11,12]. Therefore, accurate rainfall information based on a scientific study is very useful as a reference. Information relating to the future cannot be determined with certainty but can only be predicted. This study aims to predict rainfall data more precisely in Makassar city using Autoregressive Integrated Moving Average (ARIMA) and Kalman Filter methods.
2. Methods

This research focuses on modelling rainfall data in Makassar from January 2010 to December 2020. Data was collected from the website of Meteorology, Climatological, and Geophysical Agency (BMKG) Paotere station (http://dataonline.bmkg.go.id/home). The daily rainfall data are available on the website. In this research, we used the total daily rainfall data for every month so-called the amount of monthly rainfall data. The data is split into training (in-sample) (January 2010-December 2020) and testing (out-sample) (January 2021-August 2021) partitions.

Data analysis consists of two methods, namely ARIMA and Kalman Filter method. ARIMA method used the Box-Jenkins methodology which consists of four stages: identification, parameter estimation, diagnostic checking, and forecasting. At this identification stage, the first step is to determine whether the time series data is stationary or not, both in the mean and variance. The stationarity testing process on the mean can be carried out using the Augmented Dickey-Fuller test or using the visual form of a time series plot. If the data is nonstationary in the mean, the differencing process is needed. If a stationary condition in the variance is not obtained, Box & Cox (1964) introduces a power transformation $\lambda$ which is called a parameter transformation. The next steps are to plot the autocorrelation function (ACF) and the partial autocorrelation function (PACF). The order of autoregressive (AR), and moving average (MA) can be determined through both ACF and PACF plots. The next step after obtaining the tentative ARIMA ($p, d, q)(P, D, Q)$ models is to estimate the parameters of the model using the least squared method. The next step is diagnostic checking which is divided into two parts, namely the parameter significance test with the $t$-test and the model suitability test includes a test of the assumption of residuals white noise and are normally distributed. Testing whether the residual white noise or not can be used Ljung-Box test while testing whether the residuals are normally distributed or not can be used the Kolmogorov-Smirnov test. Detailed of ARIMA model is given in Part 2.1. The last stage is forecasting.

The Kalman Filter method used the Algorithm of Kalman Filter which consists of two stages namely the prediction stage and the correction stage [10,13-17]. Detailed Kalman Filter algorithm is given in Part 2.2. The best model selection is based on the smallest Mean Absolute Percentage Error (MAPE) value [18].

2.1. SARIMA (Seasonal Autoregressive Integrated Moving Average)

In general, the seasonal ARIMA Box-Jenkins, ARIMA($p, d, q)(P, D, Q$) model [19-22], is defined as follows:

$$
\phi_p(B)\Phi_p(B^s)(1-B)^d(1-B^s)^0Y_t = \theta_q(B)\Theta_q(B^s)a_t
$$

(1)

where

$$
\phi_p(B) = 1 - \phi_1B - \phi_2B^2 - ... - \phi_pB^p
$$

$$
\Phi_p(B^s) = 1 - \Phi_1BS - \Phi_2B^2S - ... - \Phi_pB^pS
$$

$$
\theta_q(B) = 1 - \theta_1B - \theta_2B^2 - ... - \theta_qB^q
$$

$$
\Theta_q(B^s) = 1 - \Theta_1B^s - \Theta_2B^{2s} - ... - \Theta_qB^{qs}
$$

$(1-B)^d$: the order of non-seasonal differencing

$(1-B^s)^0$: the order of seasonal differencing

$Y_t$: the number of observations (events) at time $t$

$a_t$: a white noise process that is assumed to have a mean of 0 and a constant variance

$(p, d, q)$: the order of autoregressive, integrated, and moving order of non-seasonal respectively

$(P, D, Q)$: the order of autoregressive, integrated, and moving order of seasonal respectively

S is seasonal
2.2. **Kalman filter**

In general, the Kalman Filter stages are grouped into two parts, namely, the prediction stage to get the value of the pre-estimated time for the next step and the measurement update stage. The measurement update equation serves for feedback purposes, such as combining the latest measurement results with pre-estimated values to get a better after-estimated value [10,15,23]. Kalman Filter model can be written in the general equation as follows:

State equation: \( \bar{X}_{t+1} = F \bar{X}_t + G \bar{a}_{t+1} \)

Output equation: \( \bar{Z}_k = H \bar{X}_t + b_t \)

where

- \( \bar{X} \) = the size of the State vector
- \( F \) = the size of the transition matrix
- \( G \) = Input matrix
- \( H \) = Matrix parameters size
- \( \bar{a} \) = Noise vector size in the state equation
- \( b \) = Noise vector size in output equation

All the steps of the Kalman Filter are given in Table 1.

### Table 1. Kalman filter stages

| Step | Equation |
|------|----------|
| **State Space** | 1. state equation \( \bar{X}_{t+1} = F \bar{X}_t + G \bar{a}_{t+1} \) |
|  | 2. output equation \( \bar{Z}_k = H \bar{X}_t + b_t \) |
| **Prediction Stage** | 1. state equation estimation \( \bar{X}_{t+1} = F \bar{X}_t + G \bar{a}_{t+1} \) |
|  | 2. covariant error \( P_{t+1} = F \Gamma_t F^\prime \) |
| **Measurement update** | 1. determine the Kalman gain \( P_{t+1} = F \Gamma_t F^\prime + G \Sigma G^\prime \) |
|  | 2. update the estimate with \( Z_k \) \( \bar{X}_{t+1} = \bar{X}_{t+1} + K_{t+1} (\bar{Z}_k - H \bar{X}_{t+1}) \) |
|  | 3. covariant error \( \Gamma_{t+1} = (1 - K_{t+1} H) \Gamma_t + K_{t+1} \Sigma K_{t+1} \) |

### 3. Results and discussion

#### 3.1. **ARIMA models**

**3.1.1. Model identification of rainfall data** The first stage in time series analysis is model identification which consists of time series plots, ACF, and PACF plots for Makassar City rainfall data [24]. The time series plot of rainfall data from January 2010 to December 2020 is given in Figure 1, while the ACF and PACF plots are given in Figure 2.

![Time series plot of rainfall data](image_url)
Figure 2. Plots of ACF and PACF rainfall data in Makassar

Figure 1 shows that the data is nonstationary in the mean. Based on the ACF plot, it can be concluded that the data is nonstationary in the mean and there is a strong variation with seasonal period 12. Because the data is nonstationary in the mean, it is necessary to do a differencing process. The first non-seasonal differencing (d=1), and the first seasonal differencing (D=1) were performed to eliminate the strong seasonal effect. Time series plot, ACF, and PACF plots after the first non-seasonal differencing with d=1 are given in Figures 3 and 4, respectively.

Figure 3. Time series plot of rainfall data after the first non-seasonal differencing with d=1

Figure 4. Plots of ACF and PACF of rainfall data after the first non-seasonal differencing with d=1

Based on the time series plot (Figure 3) and the ACF plot (Figure 4), we can conclude that the data is nonstationary on the seasonal mean (seasonal period 12). Therefore, the differencing process with seasonal differencing (D=1) was called for after the first non-seasonal differencing with d=1. Time series plot, ACF, and PACF plots after the first non-seasonal differencing (d=1), and the first seasonal differencing (D=1) are given in Figures 5 and 6 respectively.
Figure 5. Time series plot of differencing rainfall data (d=1) non-seasonal and seasonal period 12 (D=1)

Figure 6. ACF and PACF plots of differencing rainfall data (d=1) non-seasonal and seasonal period 12 (D=1)

Figure 5 indicates a stationary process with constant mean and variance. Based on Figure 6, the ACF plot has a single spike at lag 1 (cut off after lag 1) and the PACF plot cuts off after lag 2 indicate that the series is likely to be generated MA(1) process or AR(2) or ARMA (2,1) processes for non-seasonal patterns. Furthermore, based on the ACF plot of the seasonal lag, it can be seen that the ACF plot has a single spike at lag seasonal period 12 and the PACF has no spike at every lag. Hence, we identify some ARIMA (p, d, q) (P, D, Q)\textsuperscript{12} processes as tentative models for the data, namely ARIMA (0,1,1)(0,1,1)\textsuperscript{12}, ARIMA (2,1,0)(0,1,1)\textsuperscript{12}, ARIMA (2,1,1)(0,1,1)\textsuperscript{12}. We also tried other models namely ARIMA (1,1,0)(1,1,0)\textsuperscript{12} and ARIMA (1,1,1)(1,1,1)\textsuperscript{12}.

3.1.2. Parameter estimation and diagnostic checking. These tentative ARIMA (p, d, q) (P, D, Q)\textsuperscript{12} models were checked whether the parameter estimation is all statistically significant at a level of 0.05 or not. For diagnostic checking, the residuals of these tentative ARIMA (p, d, q) (P, D, Q)\textsuperscript{12} models were also tested whether the residuals are white noise and are normally distributed. Among these tentative ARIMA (p, d, q) (P, D, Q)\textsuperscript{12} models, only ARIMA (0,1,1)(0,1,1)\textsuperscript{12} fulfill these assumptions (all parameters are significant and the residuals are white noise and are normally distributed). Therefore, the best ARIMA model to forecast the rainfall data in Makassar is ARIMA (0,1,1)(0,1,1)\textsuperscript{12}. The estimation of parameters of $\theta_1$ is -0.917105, and $\Theta_1$ is -0.999984. ARIMA (0,1,1)(0,1,1)\textsuperscript{12} models is given as follows:

\[ (1 - B)^{1}(1 - B^{12})^{1}Y_t = (1 - \theta_1B)(1 - \Theta_1B^{12})a_t \]
\[ (1 - B)^{1}(1 - B^{12})^{1}Y_t = (1 + 0.91B)(1 + 0.99B^{12})a_t \]
3.1.3. Forecasting. The forecasting results are based on the best model namely ARIMA (0,1,1)(0,1,1)\textsuperscript{12}. Based on the forecasting result, the amount of rainfall data for January 2021 (675) to August 2021 (-2) decreased, but the amount of rainfall data increased from August 2021 (-2) to December 2021 (663). The amount of rainfall data on January 2021 was almost similar to the amount of rainfall in December 2021.

3.2. Kalman filter

The Kalman Filter equation is formed with the best ARIMA model (0,1,1)(0,1,1)\textsuperscript{12} which is presented in the form of state space. The Kalman Filter equation is obtained as follows:

\[
\begin{align*}
\epsilon_{t+1} & = \\
x_{t+1} & = \\
y_t & = [1 0 0 0 0 0 0 0 0 1 0 1] x_t + w_t \\
\end{align*}
\]

After all the Kalman Filter steps have been carried out, the ARIMA-Kalman Filter is predicted for the next twelve months. The comparison of the prediction results of the actual data, ARIMA and ARIMA-Kalman Filter is presented in Table 2. Table 2 shows that ARIMA-Kalman Filter prediction results are closer to actual data than ARIMA predictions. The MAPE value of the ARIMA-Kalman Filter is 147.00 which is less than the MAPE value of the ARIMA.

| Month       | Actual data | ARIMA Prediction | ARIMA-Kalman Filter prediction Results |
|-------------|-------------|------------------|---------------------------------------|
| January 2021| 702         | 675              | 684                                   |
| February 2021| 562        | 534              | 543                                   |
| March 2021  | 332         | 353              | 362                                   |
| April 2021  | 169         | 151              | 159                                   |
| May 2021    | 77          | 81               | 90                                    |
| June 2021   | 56          | 45               | 54                                    |
| July 2021   | 28          | 23               | 32                                    |
| August 2021 | 7           | -2               | 6                                     |
| September 2021| 26         | 15               | 24                                    |
| October 2021| 64          | 49               | 58                                    |
| November 2021| 169        | 151              | 159                                   |
| December 2021| 664        | 663              | 642                                   |

4. Conclusion

In conclusion, the best ARIMA model for forecasting rainfall data in Makassar, Indonesia from January 2010 to December 2020 is ARIMA (0,1,1)(0,1,1)\textsuperscript{12} with a MAPE value is 111.48. ARIMA-Kalman Filter provides more precise forecasts than the ARIMA model based on a MAPE value. These results are based on the case study of rainfall data in Makassar, Indonesia. It is acknowledged that different datasets may have different results. Therefore, investigating more generally these forecasting methods through a simulation study could be potential future work.
5. References

[1] Pemkot-Makassar 2019 *Dokumen Rencana Program Investasi Jangka Menengah: Kota Makassar 2019-2023* (Makassar: Cipta Karya)

[2] Badan Pusat Statistik Makassar 2010 *Makassar Dalam Angka 2010* (Makassar: Badan Pusat Statistika Kota Makassar)

[3] Maru R, Baharuddin I I, Umar R, Rasyid R, Uca, Sanusi W and Bayuuddin 2015 *Analysis of The Heat Island Phenomenon in Makassar, South Sulawesi, Indonesia* *Am. J. Appl. Sci.* 12

[4] Paida 2011 *Deteksi Peru bahan Pola Curah Hujan Kota Makassar* 46

[5] Ihsan H, Sanusi W and Hasriani H 2020 *Peramalan Pola Curah Hujan Di Kota Makassar Menggunakan Model Rantai Markov* *J. Math. Comput. Stat.* 2 19

[6] Kristianda F and Fithriasari K 2015 *Analysis of The Heat Island Phenomenon in Makassar, South Sulawesi, Indonesia* *Am. J. Appl. Sci.* 12

[7] BMKG Kelas IV Tana Toraja 1961 *Analisis Hujan Lebat di Sulawesi Selatan Tanggal 22 Januari 1961* 1007 55–6

[8] Giarno, Hadi M P, Suprayogi S and Murti S H 2018 *Distribution of Accuracy of TRMM Daily Rainfall in Makassar Strait* *Forum Geogr. Indones. J. Spat. Reg. Anal.* 32 38–52

[9] Rochman D F 2019 *Estimasi Parameter Model Seasonal Autoregressive Integrated Moving Average Dengan Metode Kalman Filter* *J. Sains Seni ITS* 17 92–7

[10] Kurniawan T, Hanafi L, Apriliani E, Matematika J, Matematika F and Alam P 2014 *Penerapan Metode Filter Kalman Dalam Perbaikan Hasil Prediksi Cuaca dengan Metode ARIMA.* *J. Sains dan Seni ITS*, vol. 3, no. 2, 2014, doi10.12962/j23373520.v3i2.7984

[11] Rantala J 1986 *Experience Rating of ARIMA Processes by the Kalman Filter* *ASTIN Bull.* 16 19–31

[12] Mirawati T D, Yasin H and Rusgtyonono A 2013 *Prediksi Curah Hujan dengan Metode Kalman Filter* (study kasus di Kota Semarang Tahun 2012) *J. Gaussian* 2 239–48

[13] Achmada A S 2017 *Model ARIMA-Filter Kalman untuk Prediksi Harga Komoditas Minyak Mentah* *J. Gaussian* 2 239–48

[14] Gelman A, Hwang J and Vehtari A 2014 *Understanding predictive information criteria for Bayesian models* *Stat. Comput.* 24 997–1016

[15] Paleni H 2011 *Time series Time series Time series Time series Time series Time series* *J. Gaussian* 2 239–48

[16] Aswi A and Sukarna S 2017 *Analisis Deret Waktu: Teori dan Aplikasi* ed M A Tiro (Makassar: Andira Publisher)

[17] Wei W W . 2006 *Time Series Analysis: Univariate and Multivariate Methods* (United State of America: AddisonWesley Publishing Company, Inc)

[18] Dhewaj K, Kovitz J M, Tian H, Jiang L J and Itoh T 2018 *Half-mode cavity-based planar filtering antenna with controllable transmission zeroes* *IEEE Antennas Wirel. Propag. Lett.* 17 833–6

[19] Sinay L J and Kembauw E 2021 *Monthly Rainfall Components in Ambon City: Evidence from the Serious Time Analysis* *IOP Conf. Ser. Earth Environ. Sci.* 755 012079

Acknowledgment

The authors would like to thank Dr. Aswi for her valuable suggestions.