The Ensemble Of Arima And Gstar Models In Forecasting Rainfall Using Kalman Filter

Unik Novita Wulandari, Alfian Futuhul Hadi, Kosala Dwidja Purnomo

Abstract: Several forecasting rainfalls with various models have been carried out in the same area. The results in each forecasting may be different from each other and to choose the best one is difficult. In this study we will discuss the Super-Ensemble Kalman Filter method which combines two or more forecasting results using the Kalman Filter method to get maximum results. The rainfall data used in this study has been divided into 4 clusters using K-Means. The ARIMA and GSTAR models from the 4 clusters were selected as the best model by looking at the smallest RMSE value from each model then the best of ARIMA and GSTAR models were ensembled using Kalman Filter. Based on the results obtained, the Super-Ensemble Kalman Filter method provides maximum results in forecasting rainfall data.

Index Terms: forcasting, rainfalls, ensemble*, Kalman Filter,* Super-Ensemble *Kalman Filter*

1 INTRODUCTION

Rainfall is one of the important climate factors in agriculture. Beside that drought also affects agriculture. Weather conditions in Indonesia are difficult to predict because the geographical condition that varies between mountains, beaches and lands. Therefore, a forecasting method which is in accordance with the geographical conditions in particular area is needed to minimize the farmers' loss. Forecasting is a process of estimating future conditions using data in the past. One of often used models in forecasting methods is the time series model. The time and location dependency elements combined in double variable time series are space time models and one of the multivariate models. Kalman Filter method has the advantage that forecasting can be updated with the latest data so that the predicted value is always updated [4]. Super-Ensemble Kalman Filter is basically a method that combines two or more estimated results using the Kalman Filter method in predicting rainfall. The Super-Ensemble Kalman Filter uses the same approach as a linear combination, but in this Super-Ensemble Kalman Filter, the weight and error covariance matrix move dynamically to allow better results than previous observations. In this study we used the ARIMA and GSTAR models as fitting models in the Super Ensemble Kalman Filter method The ARIMA and GSTAR models are one of the models used in rainfall forecasting. Both methods only use the basic model and the error in forecasting the rainfall. Based on that, the Super-Ensemble Kalman Filter technique is used to combine two or more forecasting results using the Kalman Filter method to get maximum results. Then we will see how the comparison between the Kalman Filter Super Ensemble and ARIMA using updated data.

- *Unik Novitta Wullandari is currently pursuing masters degree program in Mathematics Department, Faculty of Mathematics and Natural Sciences in Jember University, Indonesia, PH-085608500995. E-mail: niks_rohmah91@yahoo.com*
- *Alfian Futuhul Hadi is Lecture in Mathematics Department, Faculty of Mathematics and Natural Sciences in Jember University, Indonesia.*
- *Kosala Dwidja Purnomo is Lecture in Mathematics Department, Faculty of Mathematics and Natural Sciences in Jember University, Indonesia*

2 RESEARCH METHOD

The linear combination method describes the Super-Ensemble (SE) concepts very well.

The Super-Ensemble Kalman Filter uses the same approach as linear combination, in this filter, the weight and error covariance matrix move so dynamically that allow better results than previous observations [3]. During the study period, the formulation for forecast models is as follows:

$$\mathbf{w}_j^f = \mathbf{I}\mathbf{w}_{j-1}^a \quad , j = 1,\dots, N_I \tag{1}$$

$$\mathbf{P}_j^f = \mathbf{I}\mathbf{P}_{j-1}^a \mathbf{I}^{\mathbf{T}} + \mathbf{Q}_{j-1} , j = 1,\dots, N_I \tag{2}$$

Some elements in the diagonal $\mathbf{P}$ can be developed as observational filters in assimilating, while diagonal $\mathbf{Q}$ remain throughout the process. In the analysis step, state vectors and covariance matrices are updated by adding a prediction component that takes into account the model and uncertainty in observation, as shown in the equation

$$\mathbf{w}_j^a = \mathbf{w}_j^f + \mathbf{K}_j \left(y_j - \mathbf{x}_j \mathbf{w}_j^f \right), j = 1,\dots, N_I \tag{3}$$

$$\mathbf{P}_j^a = \mathbf{P}_j^f - \mathbf{K}_j \mathbf{x}_j \mathbf{P}_j^f \quad , j = 1,\dots, N_I \tag{4}$$

$K_j = P_j^f x_j^T \left(x_j P_j^f x_j^T + R_j \right)^{-1}$ is the Kalman gain matrix at time $-j(M \times 1)$. Finally, hindcast and forecast are computed as usual

$$h_j^{KF} = \sum_{i=1}^{M} x_{j,i} w_i \quad , j = 1,\dots, N_I \tag{5}$$

$$f_j^{KF} = \sum_{i=1}^{M} x_{j,i} w_i \quad , j = 1,\dots, N_I \tag{6}$$

In this study used rainfall data in Jember district from January 2005 to December 2017. The data from January 2005 to December 2016 were used as in-sample data, while the out-sample data used was data from January 2017 to December 2017. Those will be separated into four groups, each of it will represent a variable. Insample data is used to build ARIMA and GSTAR models while data outsample correction of forecasting results used the Super-Ensemble Kalman Filter method. In the insample data, plotting is needed. This aims to obeserve the stationarity of the data. If it is not stationary in the mean, it is necessary to have differencing process, but if the data variance is not stationary, then it is necessary to transform the data using the Box-Cox transformation. The data that has been seen for its standardity can be used to build ARIMA and GSTAR model. Those are divided into 4 clusters, so that the model will be built in each cluster. Finally we will have 4 different ARIMA and GSTAR models.The Super-Ensemble *Kalman Filter* method is the same as the Kalman Filter method. In the Super-Ensemble Kalman Filer there are also two stages, namely the stage of prediction and correction. The Rainfall data in 2016 were used as observational data to build model on the Super-Ensemble Kalman Filter. After getting the forecasting results with the Super-Ensemble Kalman Filter, the forecasting results will be compared with outsample data and forecasting results using ARIMA with updated data. The best results from the comparison can be seen from the RMSE value, where if the value of the RMSE is small, it can be concluded that the model is the best model in predicting rainfall.

3 RESULT AND DISCUSSION

In this study the data used were divided into four groups based on the grouping of BPS and grouping using the K-Means algorithm. Based on this grouping, the best forecasting model will be obtained from each group. The model will be used as a fitting model in forecasting the Super-Ensemble Kalman Filter. In the ARIMA and GSTAR models the data used in forecasting must be stationary in the mean and variance, where the data patterns do not experience significant changes [1]. Therefore, before predicting the model, it is necessary to test the mean and variance.

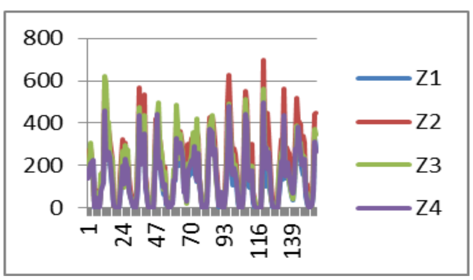

Figure 1. *Plots of rainfall data from four regions*

Figure 1 is a plot of rainfall data from the four regions in 2005-2017 with Z_1 as cluster 1, Z_2 as cluster 2, Z_3 as cluster 3 and Z_4 as cluster 4. The plot shows that rainfall in the four regions has similarities over a period of time and has a relatively constant pattern. So it can be said that all four data in the four regions have been stationary in mean and variance. Besides from looking at the stationary test graphs, it can also be seen from the results of the Augmented Dickey Fuller (ADF) to test the means of stationary means and rank test λ to test the variability of stationarity. . The best ARIMA model obtained from each cluster will be selected. The best model is ARIMA in the first cluster is ARIMA (4,0,4), the second cluster is ARIMA (3,0,1), the third cluster is ARIMA (2,0,3) and the fourth cluster is ARIMA (2,0,5). This model will be used as a fitting model in the Kalman Filter. Whereas the ARIMA Model with updated data (A2) will be used as a comparison of forecasting results with the Kalman Filter Super-Ensemble.Based on the AIC value, the best model for ARIMA (A2) in cluster 1 is ARIMA (4.0,4), Cluster 2 ARIMA (4.0,2), cluster 3 ARIMA (2,0,3) and in cluster 4 ARIMA (3,0,2). As for the GSTAR model, cluster distribution is performed using the K-Means algorithm. From Yudistira's research (2017), there are four models in each cluster. The Super-Ensemble Kalman Filter method will combine the ARIMA and GSTAR models to forecast rainfall at a certain time period in Jember Regency which has been divided into four regions. The model to be used in forecasting the Super-Ensemble Kalman Filter is assumed

$$\hat{Y} = \boldsymbol{X} \times \boldsymbol{W} \quad (7)$$

Where X is a 1×2 vector containing ARIMA and GSTAR forecasting values which are built based on 2005-2015 data per unit time (month) while W is a time vector weight which is a 2×1vector containing an initial predictive value. The Super-Ensemble Kalman Filter uses the same approach as a linear combination, but in this Super-ensemble Kalman filter, the covariance weight and error matrix move dynamically to allow better results than previous observations.

To compare the forecasting accuracy between the Super-Ensemble Kalman Filter and the ARIMA (A2) model can be done by looking at the RMSE value of each forecast based on the equation

$$RMSE = \sqrt{\frac{1}{N}\sum_{t-1}^{N}\left(Y_t - \hat{Y}_t\right)^2} \quad (8)$$

Based on forecasting results for each cluster, the RMSE values are obtained as follows

TABLE 1

RMSE OF SUPER-ENSEMBLE KALMAN FILTER AND ARIMA

Kluster	Peramalan	RMSE
1	KF	165,61
	ARIMA(A2)	267,18
2	KF	211,09
	ARIMA(A2)	563,56
3	KF	101,31
	ARIMA(A2)	263,79
4	KF	133,90
	ARIMA(A2)	165,71

Based on the RMSE values of the four clusters, the RMSE values in the Super-Ensemble Kalman Filter method are smaller when compared to the RSME values of the ARIMA (A2) model. This indicates that the Super-Ensemble Kalman Filter method is better when compared to the ARIMA (A2) model. The amount of observational data has a very important role in the accuracy of forecasting results in this method. The decrease in norm covariance error seen in the graph is a sign that the filtering process is going well. Signs and guarantees that indicate that the filtering process has been stable is the value of convergent error covariance norms [2].

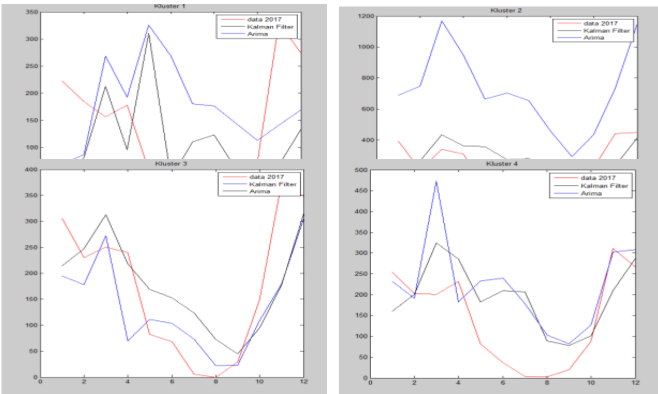

Figure 2. *Plotting of forecasting results*

4 CONCLUSION

The use of the Super Ensemble Kalman Filter method in the ARIMA and GSTAR models gives a relatively stable norm covariance error value, whereas in the Kalman Filter method a stable norm covariance error value provides good predictability results. In addition, when viewed based on RMSE values compared to the four regions in Jember Regency, rainfall forecasting using the Kalman Filter Super-Ensemble method gives better results when compared to the ARIMA (A2) model.

5 REFERENCES

[1]. Ispriyanti, D. 2004. Pemodelan Statistika dengan Transformasi Box Cox. Jurnal Matematika dan Komputer. Vol.7(3), pp: 8-17.
[2]. Kleeman, L. Tanpa Tahun. Understanding and Applying Kalman Filtering. Clayton: Departement of Electicial and Computer System Engineering MonashUniversity.http://cs.cmu.edu/~motionplanning/paper_kleeman_kal man_basics.pdf. [15 November 2018].
[3]. Lenartz, F., Beckers, J.M., Chiggiato, J., Mourre, B., Troupin, C., Vandenbulcke, L., Rixen, M. 2010. "Super-ensemble Techniques Applied to Wave Forecast: Performance and Limitations". Ocean Sci. Vol 6, pp : 595-604.
[4]. Welch, G., & Bishop, G. 2006. An Introduction to the Filter Kalman. Chapel Hill: Department of Computer Science, University of North Carolina.
[5]. Yudistira, I. 2017. Pengelompokan Stasiun Hujan Melalui Variabel Geografis pada Pemodelan GSTAR Musiman untuk Peramalan Curah Hujan di Kabupaten Jember. Skripsi. tidak diterbitkan. Jember:Universitas Jember.
[6]. Zulfi, M. 2018. Peramalan Curah Hujan di Kabupaten Jember Menggunakan ARIMA dan Kalman Filter. Tesis. tidak diterbitkan Jember:Universitas Jember.