

# IMPORTANCE OF UPDATING FOR MONTHLY RAINFALL PREDICTION BASED ON ECMWFs4

Achmad M. Rafi<sup>1\*</sup>, M. Ryan<sup>2</sup>, A. Kurniawan<sup>1</sup>, E. F. Wua<sup>1</sup>

<sup>1</sup>Malang Climatological Station, Malang, 65152

<sup>2</sup>Meteorology Climatology and Geophysics Agency, Angkasa I Street 2, Kemayoran, Jakarta Pusat, 10720, Indonesia

\*E-mail: achmadrafi08@gmail.com

Received: 30 August 2021

Revised: 14 April 2022

Accepted: 25 April 2022

## ABSTRACT

There is uniqueness in climate services in East Java. Rainfall information is delivered as N-1 analysis for ongoing months N+1, N+2, and N+3 for monthly prediction. This study aims to investigate whether updating monthly predictions improves prediction accuracy. The verification method for this study is based on the percentage accuracy of the rain class category according to SNI 8196: 2015. The data used for this study is ECMWF's monthly rainfall prediction that has three lags system (1, 2, and 3). Rasters of monthly rainfall interpolation from the main rainfall observation (197 locations) in East Java from April 2015 to May 2020 (62 months) are used for the verification process. The temporal and spatial analysis then conducted using R (+ package raster). Studies based on the local governmental zone are also used. In general, the result shows that almost all months need updating, except April-September-October. Verification of ECMWFs4 shows a better verification result (0,56) in the past five years (2016-2020) for March. The regions that need monthly updating are Bawean island, the coast of Gresik, Pasuruan, and Banyuwangi.

**Keywords:** monthly rainfall, verification, ECMWF

## 1. Introduction

To fulfil their task as a regional climate office in East Java, Malang Climatological Station has some services related to climate. One of the climate services in Malang Climatological Station is rainfall prediction. The rainfall prediction conducted in this station is a monthly prediction that has a time lag (N+1, N+2, and N+3). It means for N+1, we predict the rainfall a month ahead before the valid date. For N+2, we predict the rainfall two months before the valid date. This term also was applied for N+3, which means predicting the monthly rainfall for April in January.

In predicting rainfall, the Indonesia Agency for Meteorology Climatology and Geophysics (BMKG), as a centre of climate services office in Indonesia, uses European Centre for Medium-Range Weather Forecast (ECMWF) model. Monthly rainfall prediction is produced by ECMWFs4 and consists of 51 model members and runs for seven months ahead[1], which means when we run the model in January, we can have the monthly rainfall prediction for August. The prediction model has weaknesses due to its uncertainty in predicting the future. However, ECMWF quality is getting better and better due to the implementation of probabilistic prediction. In BMKG, ECMWF is also used to make strategic considerations to face the risk of an extreme event. Moreover, ECMWFs4 model can make long-range

precipitation predictions so that it can also be used in either predicting drought or predicting hydropower production [2-3].

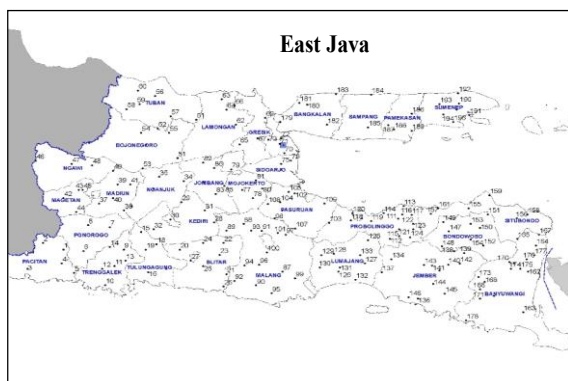
ECMWFs4 model is reliable enough for predicting rainfall anomaly over Java after got some post-processing [4]. Even though ECMWF has a satisfying performance, however, verification process still needs to be conducted. This step is essential since it's the primary process for quantifying the credibility of the climate model. In doing so, we can use statistical verification methods, such as RMSE and MAE. However, this method is too complex for the forecaster because it is hard to be understood. Moreover, the forecaster in BMKG does not use this statistical verification for their report. Instead of using the statistical method, they use percentage accuracy to verify the model result. Therefore, the gap found in these two methods needs to be bridged.

On its implementation, the forecaster has a chance to update the rainfall prediction. The ECMWF model has a time lag feature that can be used to renew the prediction output. For instance, if we have three, two- and one-time lags, it is hard to determine which model has better performance. The question of which time lag can produce a better prediction is difficult to answer. In this study, we will examine whether it is essential to update the existing rainfall prediction or if the existing prediction has already shown a good performance. Besides the temporal dimension, it is

also important to examine the spatial dimension because of the local climate variability in the study region (East Java). There are 38 local administrations in East Java, but predicting monthly rainfall is divided into 30 zones at the operational level. All city administration is merged with its district. For example, the city of Batu is joined with Malang [5]. The unique feature of each region in East Java makes the climate forecaster even harder to decide either this region need to have a monthly rainfall update with a shorter time lag or not. That is why this study aims to investigate the need for monthly rainfall updates in some of the specific regions in East Java. The detailed steps are explained in the next sections. First, data and method are presented in Sections 2. Then results and discussion are provided in Sections 3, and the last is a conclusion presented in Section 4.

## 2. Methods

The model used for this study is ECMWFs4 that is obtained from BMKG's Climatological Deputy. This model has a resolution of  $0.05^0 \times 0.05^0$  and is used to provide a monthly rainfall prediction up to 7 months ahead. One of ECMWF's outputs is comma-separated values (.csv) format. This CSV is then converted as a raster format. R Statistics software is used for processing in this study [6]. Package raster is also used to enable raster calculation [7]. In doing this study, we verify the ECMWF model result with the historical data from the 197 rainfall posts in East Java. The period of verification is from April 2015 to May 2020. This step is essential in obtaining the verification value as a result.



**Figure 1. Rainfall Observation Points in East Java (197 Points)**

**Table 1. Monthly rainfall standard categories**

Kategori	Rentang	Kategori	Rentang
1	0 - 20	6	200 – 300
2	20 – 50	7	300 – 400
3	50 – 100	8	400 – 500
4	100 – 150	9	> 500
5	150 – 200		

Data used for verification in this study is monthly rainfall observed data from 197 observation points in East Java. One hundred ninety-seven rainfall observation points are used as interpolation material for monthly rainfall information in East Java [8]. Empty data is filled using the interpolation method. In this case, Inverse Distance Weighting (IDW) is used due to its superiority compared to Spline [9]. IDW is proven as the best interpolation in East Java [10]. The parameter used in IDW is the nearest neighbour values. The interpretation of this method is one point that is influenced by a few nearest points and power. In this study, we use 12 points as the nearest neighbour parameters, and the power value equals 2, which means that the distance needs to be powered by 2. The resolution that we use is  $0.01^0$  or around 1 km with the extent of 110.79 E, 116.3 E, -9.01 S, -5 S (x-min, x-max, y-min, y-max). Raster resample then be used for providing ECMWF prediction is remain the same in both resolution and extent. The next step is removing the ocean area and the region outside East Java.

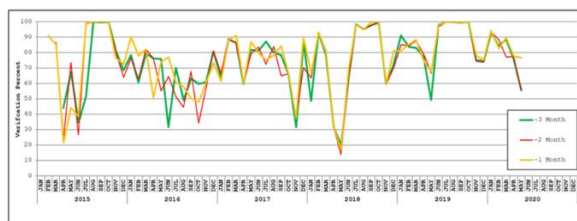
The verification method used in this study is raster-based verification. Simulation-based verification is factually based on the principle that we do not use data from the future because these data are unavailable at a specific time [11]. It means we compare raster with raster. Grid-based verification similar to the previous study is done in [12], but this study will be done with higher resolution. Higher-resolution became available because of the interpolation method. Resolution 0.01 degree can be justified based on evaluation of distance between points. From 197 points, the parameter recommends 0.018 degrees [13]. At an operational level, the main question about the prediction output is how much the prediction becoming right, so in this study, we will use binary verification. Verification used in this study is based on a contingency table [14]. The contingency table needs categories. The category used in this study is based on Indonesian National Standard (SNI) on monthly rainfall information [15]. Adjusted Proportion of Correct or called "*Sesuai Prakiraan*" is used in this study. Tolerance of single upper and/or lower categories is used. This verification is based on principle good enough information [16]. Same as operational information, we only use three lag systems which are lag-1 (N+1), lag-2 (N+2), and lag-3 (N+3). So, all verification is done in 197 points, during 62 months, and using three-time lags.

## 3. Result and Discussion

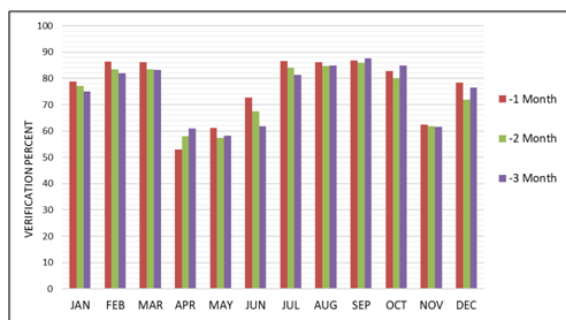
The temporal analysis of model verification and observation data have been done for April 2015 to May 2020 for using time lags of 1, 2, and 3 months. The percentage value of the verification process is displayed in Figure 2. Figure 2 shows the percentage value of the verification result. Generally, if in the lag

1 shows a higher value than in the lag 2 and lag 3, this means that the ECMWF model forecast for lag 1 can be considered as the best monthly rainfall forecast than in the lag 2 and lag 3. For instance, for the last two years (2019 and 2020), the verification result for lag 1 is higher, which means the model result for these years can be used to produce a better rainfall prediction because their result is close to the observation data. However, in 2015 and 2018, their verification result of lags-1, lag-2, and lag-3 are not satisfying, especially during the dry season. This indicates it is not recommended to use the ECMWF model result to provide the rainfall prediction.

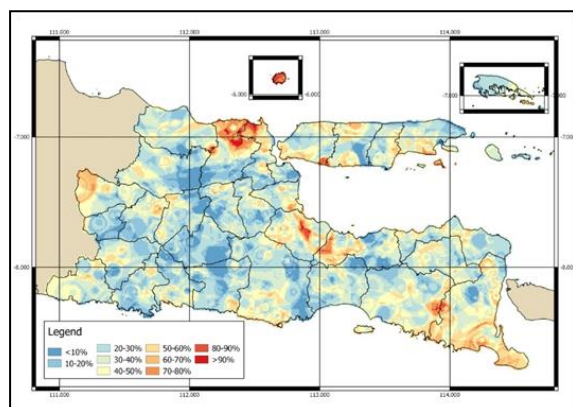
ECMWF model result is then classified into a monthly group rainfall from January to December in each lag in order to know which month is needed to be updated. The monthly verification value can be seen in Figure 3.



**Figure 2. Percentage monthly verification with time series on lag-1, lag-2, and lag-3**



**Figure 3. Percentage of monthly verification lag-1, lag-2, and lag-3**



**Figure 4. Raster of updating priority**

January has a negative slope (verification lag-1 relatively better than lag-2 or lag-3), which means that if the update is done in January, the verification will get better. The same case applies to February, March, May, June, July, August, November, and December. On the other hand, April, September, and October show a positive slope, so it is not recommended to update these three months.

If verification is deconstructed again to 62 months, as Figure 2 shows, 35 out of 62 months or around 56% need updating. It is needed for climate forecaster to downscale all these concepts to points level. But in general, the principle of negative or positive slope can guide whether prediction should be updated or not. We hope this study can illustrate how updating is done using scientific principles by creating these types of plots.

Last but not least, we want to know in spatial perspective what place updating should be done necessarily. By tightening the standard that only if verification lag-1 is right and verification lag-3 is wrong based on binary classification, thus we create a raster containing its condition. In 62 months, there will be one point if the logic is fulfilled. After we sum it on all raster, we know that the maximum point is 11. So, to make it simpler, the sum points in every single cell are divided by the maximum and multiplied by 100%. So it becomes percentage raster shows some priority of updating. 100% means it is very important to do updating in there, vice versa. Raster from the calculation then plotted using QGIS Software [17].

The higher values show higher priorities in updating (reds are the important ones). Some region in East Java such as Bawean, Gresik, Pasuruan, and Banyuwangi need updating to produce a better monthly rainfall prediction. Forecasters whose jobs in these regions need to give more attention to updating monthly rainfall prediction.

This study is far from perfect. Malang Climatological Station manages 1019 rainfall observation points in East Java. There will be more comprehensive views if we can do it at every point available. This study also does not use correction methods such as linear correction or quantile mapping. It is proven that when ECMWFs4 model output is corrected using LS and QM in Sulawesi, they will have better predictions [18]. At least this study has given some general basis and many scripts written in the R language in time further research is needed.

## 4. Conclusion

Based on the temporal and spatial verification, it can be stated that updating will bring better verification results. From a temporal perspective, at least 56%

updating brings better verification. Only verifications in April, September, and October show there was the probability of bad updating. Moreover, a raster shows updating priority is developed to reveal that Bawean, Gresik, Pasuruan, and Banyuwangi have high priority in updating monthly rainfall prediction.

## Acknowledgement

This project is one of five projects dedicated to participating in ICTMAS called ASRAB. All authors declare that there is no conflict of interest in doing this project. We would like to thanks BMKG as an institution that gives us a chance to produce these kinds of projects.

## References

- [1] Molteni F, T Stockdale, M Balmaseda, G Balsamo, R Buizza, L Ferranti, L Magnusson, K Mogensen, T Palmer dan F Vitart, *ECMWF Technical Memorandum*, 2011, 656 1-51
- [2] Carrão H, G Naumann, E Dutra, C LavayssedanP Barbosa, *Climate*, 2018, Vol. 6 (10.3390/cli6020048)
- [3] Bazile R, M A Boucher, L Perreault dan R Leconte, *Hydrol. Earth Syst. Sci.*, 2017, vol. 21, 5747–62
- [4] R Muharsyah, T W Hadidan S W Indratno, *IOP Conf. Ser.: Earth Environ. Sci.*, 2020, 572 012034
- [5] East Java's Statistics Agency 2020 Jawa Timur Province in Figures (Surabaya: BPS)
- [6] R Core Team 2020 R: “A language and environment for statistical computing” (Vienna: R Foundation)
- [7] Hijmans R J et al., “Geographic Data Analysis and Modeling, (CRAN/raster)”, 2020
- [8] Kurniawan A dan D Tarmana, *J. Met. Klim. dan Geo.*, 2019, vol. 5, 40-48
- [9] Hanif K, E Aprilia, J B Utomo, A Kurniawandan A Safril, *Pros. Sem. Nas. Geotik*, 2018, Vol 2, B Murtiyasa et al (Surakarta: University Muhammadiyah Surakarta) p 213-220
- [10] A Kurniawan, E Makmur, dan Supari, *Pros. Sem. Nas. Geomatika*, 2020, R Ati et al (Bogor: Badan Informasi Geospasial) p 263-72
- [11] R Fredyan, A Kurniawan, A Nabadan Abdurrouf 2021, *Pros. Sem. Nas. Geomatika*, 2020, R Ati et al (Bogor: Badan Informasi Geospasial) p 87-96
- [12] Ratri D N, K Whandan M Schmeits, *J. of App. Met. and Clim.*, 2019, 58 1709–23
- [13] Hengl T, *Computers & Geosciences*, 2006, vol. 32, 1283–98
- [14] Muharsyah R, *J. Met. dan Geo.*, 2017, vol. 18, 33-44
- [15] BSN-Indonesia National Standard Agency 2015 SNI 8196:2015 Spesifikasi Penyajian Curah Hujan (Jakarta: BSN)
- [16] BMKG-Indonesia Meteorology Climatology and Geophysics Agency, “Buku Verifikasi Prakiraan Iklim Indonesia”, 2013, (Jakarta: BMKG)
- [17] QGIS Development Team 2009 QGIS Geographic Information System Open Source (Geospatial Foundation): <http://qgis.org>.
- [18] Wenas GK, R Muharsyah dan S Widiyanto, *Bull. GAW Bariri*, 2020, vol. 1, 77-86