A preliminary results of assessment of BMKG-WRF numerical model daily rainfall forecasts performance using categorical verification

Fatkhuroyan, A S Praja, T Wati

Indonesia Agency for Meteorology Climatology and Geophysics BMKG, Kemayoran, Central Jakarta, Indonesia

Email: fatkhuroyan@bmkg.go.id

Abstract. Indonesia Agency for Meteorology, Climatology and Geophysics (BMKG) has been using WRF (Weather and Research Forecasting) numerical weather model in forecasting daily rainfall accumulation. The method used to compare the daily rainfall accumulation of WRF forecast results (one to three days forecast) to rain-gauge observation data from 153 meteorological stations from March 2016 to December 2017 by using categorical verification techniques. The results show that the values of Frequency Bias Index (FBI), Proportion Correct (PC), Probability of Detection (POD), False Alarm ratio (FAR), and Threat Score (TS) for one-day forecast and three-day forecast in 2016 are mostly lower than 2017. Meanwhile, for 2-days forecast in 2016 generally lower than 2017 except PC and FAR. Molucca-Papua have the highest value of PC (0.63) for one-day forecast while Kalimantan have the highest values of PC (0.67, 0.68) for 2 and 3 days ahead forecast. In contrary, Sumatera have the lowest PC value (0.58, 0.59) for 2 and 3 days ahead forecast, respectively. It can be concluded that the performance of BMKG WRF quite accurate in forecasting daily rainfall up to three days ahead.

1. Introduction
Rainfall has important role in the hydrological cycle at regional and global scales. Rainfall forecast, mainly over the tropical regions, remains a major challenge in numerical weather prediction. Rainfall with a lot of the characteristics such as high intensity, short duration, and sudden occurrence is the most complicated variable to be predicted in a mesoscale model because of its complex relationship with various forcing in the atmosphere. While the quality of predicted fields such as wind, temperature and pressure are superior to that of precipitation, rainfall is the mainly interest variable to most users and exhibits higher spatial and temporal variability. Meanwhile, large spatial variability in rainfall makes high-resolution forecasts to be essential for effective application. Moreover, when evaluate the model skills, it is very important to assess the quality of precipitation forecasts.

Atmospheric mesoscale models have appeared as promising tools for high-resolution rainfall forecasts. The numerical weather prediction (NWP) models have been successful in the last two decades for large-scale medium-range weather forecasting. However, there are several issues that need to be addressed to ensure accuracy and effectiveness of the forecasts [1]. Accurate rainfall forecast remains a challenge. Mesoscale models are widely used to obtain regional forecasts at high spatial and temporal resolution. They are forced by the initial and boundary conditions from the global model forecasts. Furthermore, They also count for the influence of detailed topography, land cover, and
vegetation, which are either missing or smoothed in global models. Several operational forecasting agencies use the mesoscale models for detailed weather forecasts over small geographical regions.

The Weather Research and Forecasting (WRF) model with the embedded The Advanced Research WRF Model (ARW) dynamical core [2] has been installed and appropriately configured in the high performance computing infrastructure of the Indonesian Agency for meteorology, climatology and geophysics (BMKG) in order to provide accurate forecasts for the entire Indonesian maritime continent.

Although the fact that the basic formulation of the mesoscale models is general, they differ from one another in many respects [3]. The skill depends on various parameterization schemes and approximations as well as initial and boundary condition [4]. It is recognized that generic mesoscale models that may be available in the public domain need an essential calibration and validation of configuration for maximum performance over a region [5,6]. Moreover, even for a given model, the skill of rainfall forecast depends on the geographical location, background state and resolution of the forecasts [7]. Globally, mesoscale forecasting is an evolving technology with many challenges; these challenges grow as the spatial resolution of the forecast is increased [8].

Many studies evaluated the performance of WRF over Southeast Asia have been conducted. [9] Evaluated statistical verification of NHM and WRF with the same horizontal resolution of 20 km and under the same conditions around Japan and Southeast Asia. [10] Attempted to verify the performance of the WRF model in downscaling the boreal summer season climate over Southeast Asia using SINTEX-F2 retrospective forecasts over the period of 2000-2013. It was found that the WRF model could improve the SINTEX-F2 forecast precipitation over Malaysia, Singapore, Brunei and western Indonesia region. [11] Verified the performance of the WRF model in reproducing the regional climate variability over South East Asia. The results indicated that the model simulates the spatial variability and annual rainfall climatology with maximum rainfall over Malaysia and Indonesia and minimum rainfall over mainland Southeast Asia and the Philippines.

In this study, the performance of the WRF weather forecasts has been assessed, using as reference the rain-gauges measurements available from BMKG. Surface observations from 153 stations were used to verify and compare categorical model forecasts of daily rainfall accumulation for 1 day ahead, for the next 2 days and for the next 3 days forecast since March 2016 until December 2017.

2. Data and methods

Daily rainfall accumulation for the NWP model were provided by BMKG. The WRF-ARW model (version 3.5) has been operationally at BMKG for several years. WRF model is initialized using NCEP GFS forecasts at 0.5° horizontal resolution. The WRF model is running operationally at a resolution of 10 km covering Indonesian Maritime Continent (figure.1). It provides daily weather forecasts for the next 120 hours. The microphysical module of the model is based on the WRF Single-Moment 5-class (WSM5) [12]. The long wave radiation scheme is RRTM scheme [13], while the Short-wave radiation scheme is Dudhia [14]. The surface layer physics is MM5 similarity based on Monin-Obukhov with Carslon-Boland viscous sub-layer [15] and the land surface model is the 4-layer unified NOAH. The parameterization of turbulence in the planetary boundary layer and in the free atmosphere follows the Yonsei University (YSU) scheme [16]. The cumulus scheme is Kain- Fritsch [17] with modification to the convection trigger.
Figure 1. The integration domain of the WRF model. The red dots indicate the location of the meteorological stations.

The comparison of the WRF rainfall forecasts against surface observations has been done across the area of the Indonesian Maritime Continent, where the available surface stations are described in Fig.1. Surface observations from almost 153 conventional stations were used to verify and compare categorical model forecasts of the daily accumulated rainfall for one day ahead, for the next 2 days and for the next 3 days from March 2016 until December 2017. Quality control has been applied to delete erroneous measurements, based on checking the physical range of each parameter being verified, the allowable rate of change in time and the stationarity.

The evaluation methodology was based on the point-to-point comparison between model-generated variables and observations. The verification scores used for the rainfall were obtained using the contingency table approach [18]. This is a two-dimensional matrix where each element counts the number of occurrences in which the model forecasts and gauge measurements exceeded or failed to achieve a certain threshold for a given forecast period. The table elements are defined as: A-if the model forecast and the gauge measurement exceeded the threshold; B-if the model forecast exceeded the threshold but the gauge measurement not; C-if the model forecast did not reach the threshold but the gauge measurement exceeded it; and D-if the model forecast and the gauge measurement did not reach the threshold, see Table 1. Considering the above elements, the Proportion Correct is defined as PC = (A+D)/N, with N holding the total number of observations being verified (N=A+B+C+D). The Probability Of Detection is defined as POD = A/(A+C), the False Alarm Ratio is defined as FAR = B/(A+B), the Frequency Bias Index is defined as FBI = (A+B)/(A+C), and Threat Score is defined as TS = A/(A+B+C).

Table 1. 2x2 contingency tabel

| Observation: Yes | Observation : No |
|------------------|------------------|
| Forecast: Yes    | A                |
| Forecast : No    | C                |
|                  | B                |
|                  | D                |

3. Results and Discussion
The accuracy of WRF prediction is quite high because the average PC value is > 0.5. The accuracy of prediction for the 3-day forecast in 2016 has the highest value of 0.79 as seen on Table 2. The accuracy of the prediction for the 1-day forecast in 2017 experienced a slight improvement compared to 2016. However, it is contrary to the 2 and 3 days forecast, the accuracy tends to be lower in 2017 than 2016. Generally, The POD values are good for all forecast days with 2017 is better than in 2016.
The TS value is also quite good although the score is almost same for all the forecast days in 2017. The error rate in the WRF prediction has a relatively low value, with FAR value reach < 0.6. Moreover, for the prediction of 1-day forecast in 2017, WRF model is able to improve the prediction in 2016, with a low FAR value of 0.48 as presented on Table 3. Meanwhile for the level of reliability based on the FBI values, the prediction of WRF tends to be over-estimated. It means that the WRF model rainfall is higher than the rain-gauge observation.

### Table 2. Categorical verification results at 2016

| Forecast     | PC  | POD | FAR | FBI | TS  |
|--------------|-----|-----|-----|-----|-----|
| 1 day ahead  | 0.58| 0.61| 0.55| 1.44| 0.34|
| 2 day ahead  | 0.74| 0.77| 0.56| 1.86| 0.39|
| 3 day ahead  | 0.79| 0.82| 0.57| 2.03| 0.39|

### Table 3. Categorical verification results at 2017

| Forecast     | PC  | POD | FAR | FBI | TS  |
|--------------|-----|-----|-----|-----|-----|
| 1 day ahead  | 0.59| 0.74| 0.48| 1.54| 0.43|
| 2 day ahead  | 0.53| 0.85| 0.53| 1.94| 0.43|
| 3 day ahead  | 0.51| 0.89| 0.54| 2.11| 0.43|

Furthermore, the discussion for all forecast days will be displayed at the next figures. The overall values of PC over Indonesia region are between 0.50-0.75 (green dots) for all the forecast day, with 2 and 3 days forecast has values > 0.75 for some places in Java, Papua and Kalimantan (blue dots). Molucca-Papua have the highest average value of PC (0.63) for one-day forecast while Kalimantan have the highest average values of PC (0.67, 0.68) for 2 and 3 days ahead forecast. In contrary, Sumatera have the lowest average value of PC (0.55) for one-day forecast while Bali-Nusa Tenggara have the lowest average value of PC (0.58, 0.59) for 2 and 3 days ahead forecast, respectively. See figure 2(a), (b) and (c). It could be because the local topography is so varied, so that the WRF model cannot predict the rainfall at those area well.
Figure 2. Overall values of PC for (a) 1-day forecast (b) 2-day forecast (c) 3-day forecast.

Figure 3(a) present overall values of POD that are > 0.5 for all 1-day forecast throughout Indonesia, except for some places in Sumatera and Java, the values are < 0.5 (red dots). Moreover, the POD values are > 0.75 for the 2 and 3-day forecast around Indonesia (blue dots) as seen on figure 3(b) and (c). POD values that are > 0.5 indicate that WRF forecasts have a fairly good accuracy in the large part of Indonesia. Rainfall forecasts for the 2 and 3 days ahead are better than 1 day ahead forecast, enabling the need for a spin up for the WRF model.
Figure 3. Overall values of POD for (a) 1-day forecast (b) 2-day forecast (c) 3-day forecast.

The values of FAR are < 0.5 (red dots) for 1-day forecast mostly all over Indonesia region as seen on figure 4(a). The values for 2 and 3-day forecast increase between 0.50 – 0.75 in some places in Sumatera, Java, Kalimantan and Papua (green dots) as seen on figure 4(b) and (c). There are a lot of FAR values < 0.5 on this forecast, It shows that the forecast has a low error rate. Only a few stations have a high error rate which is around Southeast Nusa Tenggara.
Figure 4. Overall values of FAR for (a) 1-day forecast (b) 2-day forecast (c) 3-day forecast.

The values of FBI for all forecast days generally are < 2 in mostly all over Indonesia (red dots) as seen in figure 5(a), (b) and (c). Meanwhile for 2-day forecast (figure 5(b)), the FBI values are from 2 to 4 mostly in Java, Bali, Nusa Tenggara and Maluku. Moreover, for 3-day forecast, some places in Sulawesi and Central Sumatera have values from 2 to 4 as seen on figure 5 (c). Therefore, the WRF are overestimate to the rain gauge in predicting the rainfall all over Indonesia. Rainfall forecasts for 2 and 3 days ahead are more overestimate than forecasts for 1 day ahead.
Figure 5. Overall values of FBI for (a) 1-day forecast (b) 2-day forecast (c) 3-day forecast.

The TSI values for forecast days generally from 0.25 to 0.50 for mostly throughout Indonesia (green dots) as presented in figure 6(a), (b) and (c). Meanwhile for 2-day forecast, there are some places in Sumatera and Papua have values of > 0.50 (figure 6(b)). Moreover for 3-day forecast, there are some places in Sumatera have the values >0.50 (blue dots) as seen on figure 6(c). WRF forecasts are less able to represent the actual rain conditions in the southern Nusa Tenggara region. This is evidenced by the presence of several stations in the area that have relatively small TS values compared to other regions in Indonesia.
4. Conclusions
In this study, a two-year categorical verification of rainfall was performed in the context of evaluating the WRF model forecast skill. The PC and POD values are the best at 3-day forecast over Indonesia region during all observations. Meanwhile, FAR values are the best at 1-day forecast. Moreover, the FBI values indicates that the WRF model is overestimated in predicting rainfall. Meanwhile, the TS values in 2017 are better than 2016 in some places of Sumatera and Papua at 2-day forecast. The results show that WRF model is quite good in predicting rainfall throughout Indonesia.

Acknowledgements
The authors are grateful to BMKG as Indonesian agency for Meteorology Climatology and Geophysics for the WRF model forecast and rain gauge rainfall data.

References
[1] Liu J, Bray M, Han D 2012. Sensitivity of the Weather Research and Forecasting (WRF) model to downscaling ratios and storm types in rainfall simulation. Hydrol. Processes 26, 3012–3031.
[2] Skamarock W C, Klemp J B, Dudhia, J, Gill D O, Barker D M, Dudha, M G, Huang X, Wang W, Powers Y 2008. A Description of the Advanced Research WRF Ver.30 NCAR Technical Note, NCAR/TN-475CSTR, Mesoscale and Microscale Meteorology Davison. (National Centre for Atmospheric Research, Boulder CO USA) pp 113
[3] Schmidli J, Billings B, Chow F K, de Wekker S F J, Doyle J, Grubisic V, Holt T, Jiang Q, Lundquist K A, Sheridan P, Vosper S, Whiteman D C, Wyszogrodzki A A, Zangl G 2011.
Intercomparison of mesoscale model simulations of the daytime valley wind system. *Mon. Weather Rev.* 139: 1389–1409

[4] Wen X, Lu S, Jin J 2012. Integrating remote sensing data with WRF for improved simulations of oasis effects on local weather processes over an arid region in northwestern China. *J. Hydrometeor.* 13 pp 573–587

[5] Rakesh V, Singh R, Joshi P C 2011. Evaluation of short-range forecasts from a mesoscale model over the Indian region during monsoon 2006. *Pure Appl. Geophys.* 168: 919–943

[6] Marsh P T, Kain J S, Lakshmanan V, Clark A J, Hitchens N M, Hardy J 2012. A method for calibrating deterministic forecasts of rare events. *Weather Forecast* 27: 531–538

[7] Jones C, Carvalho LM V. 2006. Changes in the activity of the Madden–Julian Oscillation during 1958–2004. *J. Climate* 19 24: 6353–6370

[8] Kotroni V, Lagouvardos K 2004. Evaluation of MM5 high-resolution real-time forecasts over the urban area of Athens, Greece. *J. Appl. Meteor.* 43: 1666–1678.

[9] Hayashi S, Aranami K, Saito K 2008. Statistical Verification of Short Term NWP by NHM and WRF-ARW with 20 km Horizontal Resolution around Japan and Southeast Asia. *SOLA* 4 pp 133–136 doi:10.2151/sola.2008-034

[10] Ratna S B, Ratnam J V, Behera S K, Doi T, Yamagata T 2016b. Downscaled prediction of extreme seasonal climate over Southeast Asia using a regional climate model. *Techno-Ocean 2016 IEEE Kobe*, pp 375–380

[11] Ratna S B, Ratnam J V, Behera S K, Tangang F T, Yamagata T 2017. Validation of the WRF regional climate model over the subregions of Southeast Asia: climatology and interannual variability. *Climate Research* 71 pp 263–280 doi: 10.3354/cr01445

[12] Hong S Y, Dudhia J, Chen S H 2004. A revised approach to ice microphysical processes for the bulk parameterization of clouds and precipitation. *Monthly Weather Review.* 132(1) pp 103–120

[13] Mlawer E J, Taubman S J, Brown P D, Iacono M J, Clough S A 1997. Radiative transfer for inhomogeneous atmosphere: RRTM, a validated correlated-k model for the long wave. *J. Geophys. Res.* 102 pp 16663–16682

[14] Dudhia, J 1989. Numerical study of convection observed during winter monsoon experiment using a mesoscale two-dimensional model, *J. Atmos. Sci.*, 46 3077–3107 doi:10.1175/1520-0469

[15] Fairall, C W, Bradley E F, Hare J E, Grachev A A, Edson J B 2003 Bulk parameterization of air-sea fluxes: Updates and verification for the COARE Algorithm. *J. Clim.*, 16: 571-591

[16] Hong S Y, Noh Y, Dudhia J 2006. A new vertical diffusion package with an explicit treatment of entrainment processes. *Monthly Weather Review* 134 pp 2318–2341, DOI: 10.1175/MWR3199.1

[17] Kain J S 2004 The Kain-Fritsch Convective Parameterization: An Update. *Journal Appl. Meteor.* 43 pp 170-181

[18] Wilks D S 2011 Statistical Methods in the Atmospheric Sciences. *3rd Edition, Academic Press, Oxford*