Improvement in WRF model prediction for heavy rain events over North Sumatra region using satellite data assimilation

Immanuel Jhonson Arizona Saragih1,2,3, Huda Abshor Mukhsinin4, Kerista Tarigan1*, Marzuki Sinambela1,2,5, Marhaposan Situmorang1, Kurnia Sembiring1, and Syahrul Humaidi1
1 Department of Physics, Faculty of Mathematics and Natural Sciences – Universitas Sumatera Utara, Medan
2 Badan Meteorologi Klimatologi dan Geofisika (BMKG), Jakarta
3 Kualanamu Meteorological Station, Deli Serdang
4 Department of Meteorology – Sekolah Tinggi Meteorologi Klimatologi dan Geofisika (STMKG), Tangerang Selatan
5 Balai Besar Meteorologi Klimatologi dan Geofisika Wilayah I, Medan

*Corresponding author: kerista@usu.ac.id

Abstract. Located adjacent to the Indian Ocean and the Malacca Strait as a source of water vapour, and traversed by the Barisan Mountains which raise the air orographically causing high diurnal convective activity over the North Sumatra region. The convective system that was formed can cause heavy rainfall over a large area. Weather Research and Forecasting (WRF) was a numerical weather model used to make objective weather forecasts. To improve the weather forecasts accuracy, especially for predict heavy rain events, needed to improve the output of the WRF model by the assimilation technique to correct the initial data. This research was conducted to compare the output of the WRF model with- and without assimilation on 17 June 2020 and 14 September 2020. Assimilation was carried out using the 3D-Var technique and warm starts mode on three assimilation schemes, i.e. DA-AMSU which used AMSU-A satellite data, DA-MHS which used MHS satellite data, and DA-BOTH which used both AMSU-A and MHS satellite data. Model output verification was carried out using the observational data (AWS, AAWS, and ARG) and GPM-IMERG data. The results showed that the satellite data assimilation corrects the WRF model initial data, so as increasing the accuracy of rainfall predictions. The DA-BOTH scheme provided the best improvement with a final weighted performance score of 0.64.

Keywords: WRF model, satellite assimilation, heavy rain

1. Introduction
Extreme weather events, such as heavy rain accompanied by strong winds, hail, lightning or thunderstorms often occur over the Indonesian Maritime Continent (IMC) which have very intensive convective cloud growth [1,2]. North Sumatra is one of the provinces on Sumatra island, located at 1°-4° N and 98°-100° E, and is geographically directly adjacent to the Indian Ocean and the Malacca Strait, also traversed by the Barisan Mountains which separating the west and the east coast of North Sumatra.
The unique topographic conditions of the North Sumatra region affect the circulation system and the weather dynamics. The existence of wide waters as a source of water vapour and mountains that raised the air orographically caused high convective activities over the North Sumatra region [3–6].

Rain events forecast as the early warning information, especially for the heavy rains, was an important component in mitigating hydrometeorological disasters. However, in BMKG operations, predictions of heavy and very heavy rain events had very low predictive skills [7]. Both strong diurnal variations and local circulation, also the rapid weather dynamics made weather forecasts in the IMC difficult to achieve the maximum accuracy [7,8].

The limitation of rainfall prediction accuracy can be overcome by utilizing the numerical weather prediction system (NWP). NWP is a predictive technique that carried out by numerically solving the general equations that form atmospheric motion in space and time [9]. Weather Research and Forecasting (WRF) is a numerical weather model that widely used for operational weather forecasting because it had high efficiency and more compatible with any computer platform [10,11].

Numerical weather model dependency on initial conditions made the limitations in the weather predictions accuracy. The fluid dynamics control equation for forecasting atmospheric conditions must have an accurate initial value to represent the real weather conditions [12]. To overcome numerical model limitations, data assimilation techniques were developed and used to correct the initial values. Data assimilation was an analysis technique in which the observed data at any predetermined time were entered into a dynamic numerical model as the initiation of initial conditions to produce atmospheric conditions forecasts as accurately as possible.

Based on the description above, this study was conducted to compare the rainfall prediction of the non-assimilated WRF model and assimilated WRF model using satellite data over the North Sumatra region. Satellite radiance data used i.e. The Advanced Microwave Sounding Unit-A (AMSU-A) and Microwave Humidity Sounder (MHS).

2. Methodology

2.1. Research time and location

This research focused on the North Sumatra region, specifically on the mainland of Sumatra Island (except Nias Island) by taking two days of heavy rain events in 2020 selected based on rainfall data from the BMKG Weather Observation Station in the North Sumatra region (Table 1), i.e. 17 June and 14 September 2020. The following are the 31 rainfall observation sites of BMKG in the North Sumatra region (Figure 1), consisting of the Automatic Agroclimate Weather Station (AAWS), Automatic Rain Gauge (ARG), and Automatic Weather Station (AWS).

| Station                                      | Code   | Latitude (°N) | Longitude (°E) | Elevation (m) |
|----------------------------------------------|--------|---------------|----------------|---------------|
| Deli Serdang Climatological Station          | 96031  | 3.62114       | 98.71485       | 25            |
| Belawan Maritime Meteorological Station      | 96033  | 3.78824       | 98.71492       | 3             |
| Kualanamu Meteorological Station             | 96035  | 3.64573       | 98.88488       | 23            |
| Deli Serdang Geophysical Station             | 96037  | 3.501         | 98.56          | 86            |
| BBMKG Region I                               | 96041  | 3.5397        | 98.64          | 0             |
| Silangit Meteorological Station              | 96043  | 2.26136       | 98.99357       | 1420          |
| Aek Godang Meteorological Station            | 96071  | 1.55          | 99.45          | 281           |
| FL Tobing Meteorological Station             | 96073  | 1.55          | 98.88          | 10            |
2.2. Data
The data used in this study are:

1) Daily rainfall observation data from weather observation stations and hourly rainfall from rainfall observation sites of BMKG in North Sumatra.

2) Global Forecast System (GFS) data, used as input data for the WRF model with the 0.25° x 0.25° spatial resolution, accessed from the NCEP GFS 0.25 Degree Global Forecast Grids Historical Archive at https://rda.ucar.edu/datasets/. Each case study used 36-hour GFS data which the first 12-hour data used as the spin-up of the WRF model.

3) The radiance data of AMSU-A and MHS satellite as WRF assimilation data on the BUFR format accessed from RDA-NCAR (Research Data Archive - National Center for Atmospheric Research) at https://rda.ucar.edu/datasets/ds735.0/.

4) Daily rainfall data of GPM IMERG (Global Precipitation Measurement - Integrated Multi-satellitE Retrievals for GPM) final precipitation data with the 0.1° x 0.1° spatial resolution accessed from NASA Earth Data at https://disc.gsfc.nasa.gov/.

2.3. WRF model configuration
This study used the Weather Research and Forecasting - Advanced Research WRF (WRF-ARW) model, which was built with the nesting and downscaling stages using two domains (Figure 2).
Figure 2. The projected domain of the WRF model used in the study; the blue box as the first domain and the red box as the second domain.

The following is the WRF model configuration scheme (Table 2), the WRFDA assimilation model experimental design (Table 3), and the 3D-Var assimilation technique using warm starts mode with a cycle interval every 6 hours (Figure 3) used in the study.

Table 2. The configuration of the WRF model

| Configuration scheme | Domain-1 | Domain-2 |
|----------------------|----------|----------|
| run-hours            | 36       |          |
| time-step            | 18       |          |
| ref_lat              | 2.8      |          |
| ref_lon              | 98.5     |          |
| geog_data_res        | 10m      | 5m       |
| Dx                   | 9 km     | 3 km     |
| Dy                   | 9 km     | 3 km     |
| e_we and e_sn        | 120      | 106      |
| e-vert               | 34       |          |
| parameterization scheme | TROPICAL physics scheme |

Table 3. The design of the WRFDA assimilation model

| Model scheme          | Description                                      |
|-----------------------|--------------------------------------------------|
| Model-1 (CTRL)        | WRF model initial GFS non-assimilated            |
| Model-2 (DA-AMSU)     | WRF model initial GFS assimilated using AMSU-A data |
| Model-3 (DA-MHS)      | WRF model initial GFS data assimilated using MHS data |
| Model-4 (DA-BOTH)     | WRF model initial GFS data assimilated using both AMSU-A and MHS data |
2.4. Validation method
The validation was carried out to compare between results of the rainfall prediction of the WRF model with and without assimilation and the observational data. The validation of the rainfall prediction output from the WRF model was carried out using standard statistical methods in two ways. The first, time-series validation compared the hourly rainfall between observation data from 31 rainfall observation sites and the simulated model values in the nearest grid element. Second, spatial validation which compared with the IMERG GPM rainfall data. The data resolution of both the WRF model and GPM IMERG reset first using the biliary interpolation method through the Climate Data Operators (CDO) application to make the same resolution.

The following are the equations of the basic statistical methods used in validation techniques of this study [13,14]:

\[
\text{Mean error (bias)} \quad BIAS = \frac{1}{N} \sum_{i=1}^{N} (P_i - O_i) = m_p - m_o
\]

(1)

\[
\text{Root mean square error (RMSE)} \quad RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (P_i - O_i)^2}
\]

(2)

\[
\text{Coefficient of determination} \quad R^2 = \left( \frac{\sum_{i=1}^{N} (O_i m_o)(P_i m_p)}{N \sigma_o \sigma_p} \right)^2
\]

(3)

where \( P_i \) and \( O_i \) as the forecast and observation values with the amount of data as much \( N \), \( m_p \) and \( m_o \) as the average of forecast and observation values, \( \sigma_p \) and \( \sigma_o \) as the standard deviation of the forecast and observation values.

3. Results and discussions
The following is the accumulated daily rainfall on 17 June and 14 September 2020 based on rainfall observation data from 6 weather observation stations in the North Sumatra region (Table 4).

| Date              | Daily rainfall accumulation (mm.day\(^{-1}\)) |
|-------------------|---------------------------------------------|
|                   | 96031 | 96033 | 96035 | 96037 | 96041 | 96043 | 96071 | 96073 |
| 17 June 2020      | 50.3  | 80.5  | 40.1  | 47.6  | 95.1  | 21.5  | 18.2  | 1.8   |
| 14 September 2020 | 64.9  | 56.1  | 70.1  | 23.4  | 27.2  | 1.0   | 0.1   | 0.8   |
3.1. Case study on 17 June 2020

The following are the results of point validation (Figure 4) and spatial validation (Figure 5) using simple statistical methods on 17 June 2020.

**Figure 4.** Box and whisker plot of mean bias, coefficient of determination, and RMSE of the time-series rainfall validation on 17 June 2020

**Figure 5.** Graph of average values of coefficient of determination ($R^2$), mean error (%), and RMSE of spatial rainfall validation on 17 June 2020

Based on the box and whisker plot graph (Figure 4), it is known that the satellite data assimilation of the WRF model increased the performance quantitatively of rainfall predictions on 17 June 2020. This was indicated by the fluctuating distribution of the statistic validation result. Based on the median value of each parameter, it is known that the lowest mean error value is the DA-BOTH scheme, the highest coefficient of determination value is the DA-MHS scheme, and the lowest RMSE value is the DA-AMSU scheme. Spatially in Figure 5, it can be seen that the coefficient of determination value in the assimilation and non-assimilation schemes are very weak, about 0.085-0.124, with the highest value being the DA-AMSU scheme. The DA-MHS scheme has the lowest RMSE value with a small difference from other schemes. Besides, all schemes forecast result was overestimated, with the best value is the DA-BOTH scheme.
The following is the distribution of the accumulated rainfall (Figure 6) and the distribution of bias error in the accumulated rainfall (Figure 7) on 17 June 2020.

**Figure 6.** Distribution of daily rainfall accumulation (mm.day\(^{-1}\)) on 17 June 2020; (a) GPM-IMERG, (b) CTRL, (c) DA-AMSU, (d) DA-MHS, (e) DA-BOTH

**Figure 7.** Distribution of bias error (%) of the daily rainfall accumulation on 17 June 2020; (a) CTRL, (b) DA-AMSU, (c) DA-MHS, (d) DA-BOTH

The spatial rainfall distribution of IMERG GPM (Figure 6.a) shows that the heavy rainfall (rainfall > 50 mm.day\(^{-1}\)) spreading across the east coast of North Sumatra. The rainfall accumulation predicted by each assimilation scheme shows a relatively similar pattern, i.e. the heavy rainfall areas along the east coast of North Sumatra. On the bias error distribution (Figure 7), it can be seen that the four assimilation schemes show a relatively similar pattern, dominated by underestimated forecasts.
(represented in blue areas). The DA-BOTH scheme has a rainy area pattern that is most suitable for GPM IMERG compared to DA-AMSU and DA-MHS. Overestimate prediction scattered in the eastern, south-eastern and north-western areas of North Sumatra.

3.2. Case study on 14 September 2020

The following are the results of point validation (Figure 4) and spatial validation (Figure 5) using simple statistical methods on 14 September 2020.

Based on the box and whisker plot graph (Figure 9), it is known that the satellite data assimilation of the WRF model increased the performance quantitatively of rainfall predictions on 14 September 2020. This was indicated by the fluctuating distribution of the statistic validation results, the assimilation schemes have higher values than the CTRL scheme. Based on the median value of each parameter, it is known that the lowest mean error value is DA-BOTH and DA-MHS schemes, the highest coefficient of determination value is the DA-AMSU scheme, and the lowest RMSE value is DA-BOTH. Spatially in Figure 9, it can be seen that the coefficient of determination value in the assimilation and non-assimilation models are very weak, about 0.013-0.022, with the highest value being the DA-BOTH.
scheme, the DA-MHS scheme having the lowest RMSE value with a small difference from other schemes. Generally, all models show overestimate results, with the best value is the DA-AMSU scheme.

The following is the distribution of the accumulated rainfall (Figure 10) and the distribution of bias error in the accumulated rainfall (Figure 11) on 14 September 2020.

**Figure 10.** Distribution of daily rainfall accumulation (mm.day$^{-1}$) on 14 September 2020; (a) GPM-IMERG, (b) CTRL, (c) DA-AMSU, (d) DA-MHS, (e) DA-BOTH

**Figure 11.** Distribution of bias error (%) of the daily rainfall accumulation on 14 September 2020; (a) CTRL, (b) DA-AMSU, (c) DA-MHS, (d) DA-BOTH

The spatial map of IMERG GPM rainfall (Figure 11.a) shows that heavy rainfall (rainfall > 50 mm.day$^{-1}$) was concentrated in the northern part of the east coast of North Sumatra. Rainfall accumulation predicted by each assimilation scheme shows a relatively similar pattern, heavy rain areas focused on the northern and southern areas of North Sumatra. By using the eye-ball verification
technique, it is shown that DA-BOTH has the most suitable rainy area pattern with GPM IMERG compared to DA-AMSU and DA-MHS. On the bias error distribution map (Figure 12), it can be seen that the four assimilation schemes show a relatively similar pattern, dominated by the underestimated forecast (represented in blue areas), and the overestimate forecast concentrated in the southern areas of North Sumatra.

4. Conclusions
Based on the verification conducted on the prediction of the non-assimilated and assimilated WRF models, it was found that the overall assimilation of the WRF model using AMSU-A and MHS satellite data performed better than the non-assimilation. The output accuracy rate of the WRF model with satellite data assimilation is higher than that of the WRF model without assimilation. This shows that assimilation provides an increase in the quantitative precipitation forecast (QPF), although the increase is not very significant. Satellite non-assimilation and assimilation schemes still tend to underestimate rainfall predictions. The DA-BOTH scheme provides a greater improvement than the DA-AMSU and DA-MHS for predicting rain events in the North Sumatra region, with a final weighted performance score of 0.64.

The high frequency of heavy rainfall occurrences in the North Sumatra region needed further research to produce a WRF model output with good accuracy. Further research is needed on the effect of satellite data assimilation on the WRF model in more case studies. Besides, it is needed to make a background error that is suitable for a long period so that the calculated error covariance has a better statistic to minimize the errors generated by the WRF model.

References
[1] Tjasyono B H 2012 Meteorologi Indonesia Volume I vol I, ed S Suratno, W Fitria and D L Sari (Jakarta: BMKG)
[2] Tjasyono B H and Harijono S W B 2012 Meteorologi Indonesia Volume II Awan dan Hujan Monsun (Jakarta: BMKG)
[3] Wirjohamidjojo S and Swarinoto Y 2010 Iklim Kawasan Indonesia (Dari Aspek Dinamik - Sinoptik)
[4] Saragih I J A, Putra A W, Nugraheni I R, Rinaldy N and Yonas B W 2017 Identification of the Sea-Land Breeze Event and Influence to the Convective Activities on the Coast of Deli Serdang IOP Conf. Ser. Earth Environ. Sci. 98
[5] Paski J A I, Saragih I J A, Permana D S, Hastuti M I, Kristianto A and Makmur E E S 2019 Simulation of land-sea breeze effect on the diurnal cycle of convective activity in the eastern coast of north Sumatra using WRF model AGERS 2019 - 2nd IEEE Asia-Pacific Conf. Geosci. Electron. Remote Sens. Technol. Underst. Forecast. Dyn. Land, Ocean Marit. Proceeding 67–71
[6] Gustari I 2009 Analisis Curah Hujan Pantai Barat Sumatera Bagian Utara Periode 1994-2007 J. Meteorol. dan Geofis. 10 29–38
[7] Gustari I, Hadi T W, Hadi S and Renggono F 2012 Akurasi prediksi curah hujan harian operasional di Jabodetabek: Perbandingan dengan model WRF J. Meteorol. dan Geofis. 13 119–30
[8] Tjasyono B H K and Harijono S W B 2017 Meteorologi Indonesia 2: Awan & Hujan Monsun (Jakarta: BMKG)
[9] Lynch P 2006 The Emergence of Numerical Weather Prediction (Cambridge University Press)
[10] Skamarock W C, Klemp J B, Dudhia J, Gill D O, Zhiquan L, Berner J, Wang W, Powers J G, Duda M G, Barker D M and Huang X-Y 2019 A Description of the Advanced Research WRF Model Version 4 (Colorado: NCAR)
[11] Skamarock W C, Klemp J B, Dudhia J, Gill D O, Barker D M, Duda M G, Huang X-Y, Wang W and Powers J G 2008 A Description of the Advanced Research WRF Version 3 (Colorado: NCAR)
[12] Barker D, Huang W, Guo Y-R and Bourgeois A 2003 A three-dimensional variational (3DVAR) data assimilation system for use with MM5 NCAR Tech Note 68
[13] Wilks D S 2016 ‘The stippling shows statistically significant grid points’: How research results are routinely overstated and overinterpreted, and what to do about it Bull. Am. Meteorol. Soc. 97
[14] Beaumont G P 2012 Intermediate mathematical statistics (Springer Science)