Spatial fraction verification of high resolution model

I Made Kembar Tirtanegara* and Fitria Puspita Sari

1Department of Meteorology, College of Meteorology, Climatology and Geophysics, Pondok Aren, South Tangerang, Banten 15521
kembartirtanegara@gmail.com

Abstract. Fraction Skill Score (FSS) is one of spatial verification method to evaluate model performance on spatial scale variations. The method was applied to assess the Weather Research and Forecasting (WRF) model using 2 km (MODEL2KM) and 6 km (MODEL6KM) grid size. Cloud Top Temperature (CTT) data from Himawari-8 satellite was utilised as a ground truth data. This study aims to evaluate the model performance by FSS with absolute and percentile threshold on convective cloud simulation for three heavy rain events. The threshold considers the evaluation of absolute and percentile aspect. The result shows that there is no significant change in the FSS value for resolution increase of MODEL2KM compared to MODEL6KM. Also, the events of heavy rain having a lower CTT generate a higher FSS value for absolute threshold. Whilst, the percentile threshold for three cases have a greater FSS value, though it cannot provide the information of CTT absolute temperature value.

Keyword: Spatial verification, FSS, WRF, could top temperature, Himawari-8

1. Introduction

The Numerical Weather Prediction Model (PCN) is a tool designed to produce fast and accurate weather forecasts [1]. With advances in computation, numerical weather prediction can simulate weather phenomena in high spatial resolution [2]. For forecasters high-resolution spatial forecasts are very useful because they present data in a more realistic grid and contour form. To determine the benefits of spatial forecasting, a precise verification method is needed that can accommodate very complex spatial structures [3].

Forecast verification is an activity to determine the relationship between forecast results and observational data [4]. The verification methods can be grouped into 4 groups, namely visual (eyeball), dichotomy, categorical, and spatial [5]. According to Mittermaier et al. [6] traditional verifications such as dichotomous and categorical methods calculated from contingency tables are not optimal in assessing the results of forecast models. Traditional verification scores with dichotomous and categorical methods provide poor information about the quality of forecasts because they only make point-based calculations without paying attention to spatial information [3]. While the visual method or eyeball, although very easy to apply, has a very subjective weakness in assessing a prediction.

Spatial verification is verification of forecasts that are spatial in nature. Spatial verification is best used in numerical weather forecasting models (PCN) because the forecast results are in the form of grids or contours [7]. New spatial verification techniques can be broadly grouped into four categories, namely neighborhood or fuzzy, scale separation or scale decomposition, features-based or object-based, and finally field deformation [8]. Although it is more complex than traditional methods, spatial verification methods can provide more objective results in the forecast verification process [9]. From various neighborhood methods, the Fraction Skill Score (FSS) verification method is widely used in weather modeling centers to evaluate high-resolution forecasts [10].

Roberts and Lean [11] introduced the FSS spatial verification method for assessing high-resolution rainfall forecasts. FSS verification is a spatial verification method that compares the fraction of model forecasts and observations in spatial scale variations with a certain threshold. Mittermaier et al. [6] spatial verification of FSS has the advantage of minimizing the occurrence of miss and false alarms on
adjacent grids in high-resolution forecasting models. FSS considers good forecasts, namely those that have spatial closeness and similarity in the fraction of events to observations [12].

Fraction Skill Score (FSS) verification is a verification method that compares forecast and observation data that have been projected in the same area coverage and resolution. Each forecast and observation grid counts the number of grids in a certain area (spatial scale) that exceeds a predetermined threshold. To calculate the FSS value, the threshold is the first thing that is applied in the process of verifying the forecast. The threshold in FSS verification consists of two types, namely the absolute threshold and the percentile [11]. Each scheme has a function and purpose that depends on the event that you want to verify. Selection of the correct threshold can reduce bias and focus on spatial accuracy [13].

Based on the background and previous research, the author tries to use FSS verification to verify the results of the Weather Research and Forecasting (WRF) spatial forecasts. To show the performance of the verification method, the author tries to verify the FSS with a different threshold scheme in the WRF spatial forecast during the convective cloud event of heavy rain. This research is expected to be able to provide objective verification of the WRF model in identifying the phenomenon of heavy rain convective clouds.

2. Experimental Design
2.1. The model and forecast

![Figure 1. The domain of research area](image)

The case study was taken from the reporting of heavy rain events on January 31 2018, March 5 2018, and March 13 2018. The timing of convective cloud events in terms of the highest rainfall on the day of the heavy rain event. In this study, the model was integrated using GFS data for 36 hours, where the first 12 hours were the spin up period of the model. The spin up time of 12 hours is the time required for the model to approach normal balance in predicting the weather [14]. The horizontal grid resolution in the first domain is 18000 m, the second domain is 6000 m, and the third domain is 2000 m. The image of the domain used in this study is shown in Figure 1. The model configuration using the tropical parameterization scheme is shown in Table 1. Tropical parameterization was chosen because since WRF 3.9 NCAR has provided a parameterization package based on latitude in the WRF model, one of which is tropical suitable for tropical areas [15].
Table 1. Configuration WRF model

| Configuration                      | Domain 1          | Domain 2          | Domain 3          |
|-----------------------------------|-------------------|-------------------|-------------------|
| Horizontal grid resolution        | 4500 m            | 1500 m            | 500 m             |
| Temporal resolution               | 60 minute         | 60 minute         | 60 minute         |
| Grid\textsubscript{west-east}     | 150               | 136               | 124               |
| Grid\textsubscript{south-north}   | 150               | 136               | 124               |
| Grid\textsubscript{vertical}      | 32                | 32                | 32                |
| Time step                         | 90 s              | 90 s              | 90 s              |
| Microphysics schemes              | WSM6 (WRF Single-moment 6 class) | WSM6 (WRF Single-moment 6 class) | WSM6 (WRF Single-moment 6 class) |
| Cumulus parameterization scheme   | New Tiedtke       | -                 | -                 |
| PBL Scheme                        | YSU (Yonsei University Scheme) | YSU (Yonsei University Scheme) | YSU (Yonsei University Scheme) |
| Shortwave radiation scheme        | RRTMG (Rapid Radiative Transfer Model for GCMs) | RRTMG (Rapid Radiative Transfer Model for GCMs) | RRTMG (Rapid Radiative Transfer Model for GCMs) |
| Longwave radiation scheme         | RRTMG (Rapid Radiative Transfer Model for GCMs) | RRTMG (Rapid Radiative Transfer Model for GCMs) | RRTMG (Rapid Radiative Transfer Model for GCMs) |

2.2. Verification method
This study uses the Fraction Scill Score (FSS) verification method using the Himawari-8 satellite CTT data as a verifier. FSS verification requires data that has the same resolution and spatial extent. First, by using the CDO application, the conversion process is carried out to NetCDF. Then the resolution of the satellite data and the output of the experimental model is equated as in Table 2 with reference to the resolution domains of two and three WRF models. The data used is IR Himawari-8 band 13 data with NetCDF format. The data is obtained from the BMKG Satellite Image Center. The satellite data has a temporal resolution of 10 minutes and a spatial resolution of 2000 meters.
Table 2. Verification experiment.

| VERIFICATION | MODEL (GRID SIZE) | SATELLITE (GRID SIZE) |
|--------------|-------------------|-----------------------|
| FSS1         | MODEL2KM (2 km)   | HIMAWARI-8 (2 km)     |
| FSS2         | MODEL6KM (6 km)   | HIMAWARI-8 (6 km)     |

This verification uses two thresholds, namely the absolute threshold and the percentile. FSS verification with absolute thresholds uses three thresholds, namely -15, -30, and -50 °C. Threshold -30 and -50 °C refer to research by Houze [16] which is interpreted to be the temperature limit for the cloud tops of Mesoscale Convective Systems. Meanwhile, FSS verification with the percentile threshold consists of 60th, 75th, and 95th. The 75th threshold describes the widespread area, namely 25% of the highest score and the 95th describes the local area, namely 5% of the highest score [11]. The FSS value is determined by the fraction value in the binary field from the satellite data and model forecasts. The fraction is calculated by dividing the number of grids in the binary plane which is 1 and the total number of grids (N) on the specified spatial scale as in equation 1.

$$FSS = 1 - \frac{\frac{1}{N} \sum_{j=1}^{N} (P_{fcst} - P_{obs})^2}{\frac{1}{N} \left( \sum_{j=1}^{N} P_{fcst}^2 + \sum_{j=1}^{N} P_{obs}^2 \right)}$$  (1)

3. Result and discussion
3.1. Cloud Top Temperature (CTT) distribution
The spatial distribution of Cloud Top Temperature is used to determine the distribution of data and to determine the absolute threshold in the simulated case. The results of the forecast of three cases are displayed to determine the comparison between the Fraction Skill Score (FSS) curve and the visual interpretation of forecasting skills. This comparison was carried out by visualizing the MODEL2KM, MODEL6KM, and Himawari-8 satellite observation data. The selected case study is the heavy rain that occurred at the Lombok Meteorological Station. The date and time are selected according to the reporting period of the highest rainfall, so that a convective cloud with a CTT of less than -30 °C is obtained.

The first case study is a heavy rain event which occurred on 31 January 2018 at 21.00 UTC. The forecast results of MODEL2KM and MODEL6KM show almost the same forecasts as shown in Figure 2. The forecast model has a convective cloud pattern that is almost the same as the HIMAWARI-8 observation data. In this case, the convective cloud area is very large and almost covers the entire research domain. The Himawari-8 observation data has a minimum CTT value of only -81.1 °C with a maximum value of 12 °C. Meanwhile, the CTT values of MODEL2KM and MODEL6KM have a minimum cloud peak temperature of -80.2 °C and -80.3 °C.
Figure 2. Cloud Top Temperature (CTT) 31 January 2018 21.00 UTC a) HIMAWARI-8, b) MODEL6KM, dan c) MODEL2KM

In the event of March 5 2018 at 09.00 UTC (Figure 3), it can be seen that MODEL2KM and MODEL6KM also produce almost the same forecast. The Himawari-8 observation data has a CTT value range of -70 °C to 21 °C. Meanwhile, the CTT values for MODEL2KM and MODEL6KM had minimum cloud peak temperatures of -82.6 °C and -81.8 °C. The minimum value of the forecast model's CTT has a value that is close to the surface temperature due to the absence of clouds in the area. Spatially, the model has a convective cloud pattern that is almost the same as HIMAWARI-8. However, the model has a CTT area that is less than -30 °C wider than the HIMAWARI-8.

Figure 3. Cloud Top Temperature (CTT) 5 March 2018 09.00 UTC a) HIMAWARI-8, b) MODEL6KM, dan c) MODEL2KM
On March 13 2018 at 09.00 UTC (Figure 4), there was no significant difference between MODEL2KM and MODEL6KM. The Himawari-8 observation data has a minimum CTT value of -63.7 ºC. Meanwhile, the minimum cloud top temperatures of MODEL2KM and MODEL6KM are -82.6 ºC and -81.2 ºC, respectively. Thus, in the case study of 13 March 2018, the two model results had the largest difference in minimum CTT among other case studies. Spatially, the two models also have different and wider convective cloud patterns and areas than HIMAWARI-8.

Visualy (eyeball), the CTT forecast results on MODEL6KM and MODEL2KM have almost the same results. However, the minimum value of the HIMAWARI-8 cloud peak temperature is always higher than the predicted cloud top temperature. The comparison of the forecast model with the HIMAWARI-8 CTT has mixed results. Of the three cases analyzed by eyeball, the case study on 31 January 2018 showed the results of the model that were closest to the CTT value and cloud patterns with temperatures less than -30º C. Meanwhile, the case study 13 March 2018 showed the results that were at least in accordance with HIMAWARI's observation data. So, the next section is carried out further verification to determine the quantification of the suitability between the model and the HIMAWARI-8 observation.

3.2. Absolute threshold verification
FSS verification with absolute thresholds uses three thresholds of -15, -30, and -50 ºC. The thresholds of -30 and -50 ºC refer to the research of Houze [16] which is interpreted to be the temperature limit for the cloud top temperature of Mesoscale Convective Systems. The absolute threshold is also adjusted to the visual results from the previous section, where this threshold does not exceed the minimum value of the CTT model or HIMAWARI-8. Meanwhile, the neighborhood scale starts from 6-246 km so that the FSS MODEL2KM and MODEL6KM charts have the same neighborhood scale.

Figure 4. Cloud Top Temperature (CTT) 15 March 2018 09.00 UTC a) HIMAWARI-8, b) MODEL6KM, dan c) MODEL2KM
In Figure 5, it can be seen that MODEL2KM and MODEL6KM have almost the same score. On 31 January 2018 the model produced excellent forecasting skills. This is because the FSS value of all thresholds has reached a value of 0.9 on the smallest neighborhood scale and is closer to 1 on the larger neighborhood scale (Figure 5a). Furthermore, on the 5 March 2018 incident, MODEL2KM and MODEL6KM also showed good skills, but only at the threshold of -15 and -30 with FSSuniform being reached in the 6 km neighborhood. Meanwhile, at the threshold of -50 °C the new model reaches FSSuniform on a neighborhood 60 scale as shown in Figure 5b. In contrast to 5 March 2018, the FSS chart in the case of 13 March 2018 shows poor skills where skills for a threshold of -30 and -50 °C have a very low FSS value ranging from 0.2 to 0.42 while at a threshold of -15°C the FSS value ranges from 0.55, up to 0.78. This shows that in the event of 13 March 2018 the forecast model could not detect convective clouds with a limit of -30°C and 50°C (Figure 5c).

The graphic results of the three cases in Figure 5 show that MODEL2KM has a higher FSS value than MODEL6KM in the case of 13 March 2018. Judging from the CTT distribution in Figure 4, it will show that MODEL2KM gets a better FSS value when the CTT distribution is uneven or present. many cloud cells in the verification domain. Of the three cases verified by FSS, the absolute threshold shows that at the time of 31 January 2018 the model had very good prediction skills. When compared with the eyeball analysis on the 31 January 2018 case, it is also the simulation model with the best performance.

3.3. Percentile threshold verification
In this section, the FSS value is calculated based on the percentile threshold of 60th, 75th, and 95th spatially (neighborhood) and temporally (time series). The 75th threshold describes a widespread area of 25% of the highest value Cloud Top Temperature (CTT) and the 95th describes the local area, namely 5% of the highest value [11]. The 75th threshold of the CTT value also indicates a large convective cloud area, while the 95th indicates the cloud top area.
Table 3. Percentile value for forecast model and HIMAWARI-8

|                  | 31 January 2018 | 5 March 2018 | 13 March 2018 |
|------------------|----------------|--------------|---------------|
|                  | MODEL          | HIMA-8       | MODEL         | HIMA-8        | MODEL         | HIMA-8        |
| 60<sup>th</sup>  | -62.3 °C       | -67.3 °C     | -61.2 °C      | -14.6 °C      | -48.4 °C      | -0.7 °C       |
| 75<sup>th</sup>  | -69.2 °C       | -70.7 °C     | -69.2 °C      | -38.4 °C      | -58.1 °C      | -11.7 °C      |
| 95<sup>th</sup>  | -76.5 °C       | -75.5 °C     | -76.5 °C      | -59.8 °C      | -68.2 °C      | -45.6 °C      |

Table 3 shows the percentile values for the HIMAWARI-8 model and observation data which is used as the threshold. The percentile values of the model and HIMAWARI-8 produce completely different absolute values, except on 31 January 2018. The difference between the model and HIMAWARI-8 percentile values is considered to eliminate bias due to differences in absolute values. From this value, it can be seen that the model value is close to the observation, namely 31 January 2018. However, this does not guarantee a high FSS value because it depends on the spatial grid being assessed.

Figure 6. Fraction Skill Score (FSS) graph with percentile threshold a) 31 January 2018, b) 5 March 2018, and c) 13 March 2018

In Figure 6, it can be seen that MODEL2KM and MODEL6KM have almost the same FSS value. 5 March 2018 (Figure 6 b) shows good skills because the 60<sup>th</sup> and 75<sup>th</sup> FSSuniform threshold can be reached in the 6 km neighborhood. Meanwhile, the 95<sup>th</sup> FSSuniform threshold was reached in the 100 km neighborhood. Meanwhile, 13 March 2018 showed excellent skills where FSS and FSSuniform had been achieved on a neighborhood scale of less than 50 km. Unlike the others, the forecast model on 31 January 2018 shows poor skills, especially at the 95<sup>th</sup> threshold which indicates that the cloud center temperature is not spatially close to HIMAWARI-8.
4. Conclusion

- Increasing the resolution through MODEL2KM does not get a significant change in the FSS value compared to MODEL6KM.
- The accuracy of the forecast model for convective cloud events with FSS verification is highly dependent on the simulated CTT distribution. Thus, the use of an absolute threshold in the case of a convective cloud that has a lower CTT and is wider has a higher FSS value. Meanwhile, the use of the percentile threshold in cases with a distribution pattern that matches the observation has a better FSS value. However, the use of the percentile threshold cannot provide information on the absolute temperature value of the CTT.

5. Acknowledgments
The authors wish to thank to the Meteorology, Climatology and Geophysics Agency (BMKG) and the Indonesian Institute of Sciences (LIPI) for facilitating in this research

References
[1] Kalnay E 2003 Atmospheric Modelling, Data Assimilation, and Predictability Cambridge University Press United Kingdom.
[2] Done J Davis C.A and Weisman M 2004 The Next Generation of NWP: Explicit Forecasts of Convection Using The Weather Research and Forecasting (WRF) Model Atmospheric Science Letters 5.
[3] Casati B 2010 New Developments of The Intensity-Scale Technique within The Spatial Verification Methods Intercomparison Project Weather and Forecasting 25.
[4] World Meteorological Organization 2000 WMO No. 1023 Guidelines on Performance Assessment of Public Weather Services Secretariat of the World Meteorological Organization.
[5] JWGFVR 2017 Forecast verification: issues, methods and FAQ http://www.cawcr.gov.au/projects/verification
[6] Mittermaier M Roberts N and Thompson S A 2013 A Long-term Assessment of Precipitation Forecast Skill Using the Fractions Skill Score Meteorological Applications 20 176 – 186
[7] Gilleland E Ahijevych D Brown B G Ebert E 2010 Ferifiying Forecast Spatially Bulletin of the American Meteorological Society 91 1365 – 1376.
[8] Gilleland E Ahijevych D Brown B G Casati B and Ebert E E 2009 Intercomparison of Spatial Forecast Verification Methods Weather and Forecasting 24 1416 – 1430.
[9] Ahijevych D Gilleland E Brown B G and Ebert E E 2009 Application of Spatial Verification Methods to Idealized and NWP-Gridded Precipitation Forecasts Weather and Forecasting 24 1485 – 1497.
[10] Ebert E Wilson L Weigel A Mittermaier M Nurmi P Gill P Gober S Joslyn B Brown T Fowlerh and Watkins A 2013 Progress and Challenges in Forecast Verification Meteorological Applications 20 130 – 139.
[11] Roberts N M and Lean H W 2008 Scale-Selective Verification of Rainfall Accumulations from High-Resolution Forecasts of Convective Events Monthly Weather Review 136 78 – 97
[12] Ebert E E 2009 Neighborhood Verification: A Strategy for Rewarding Close Forecasts Weather and Forecasting 24
[13] Mittermaier M and Roberts N 2010 Intercomparison of Spatial Forecast Verification Methods: Identifying Skillful Spatial Scales Using the Fractions Skill Score Weather and Forecasting 25 343 – 354.
[14] Rahman M M Lu M and Kyi K H 2016 Seasonality of Hydrological Model Spin-Up Time: a Case Study Using the Xinanjiang Model Journal Hydrology and Earth System Sciences
[15] NCAR 2017, More Runtime Options, WRF Tutorial, UNSW, Australia
[16] Houze R. A 2018 100 Years of Research on Mesoscale Convective Systems