Verification of multiresolution model forecasts of heavy rainfall events from 23 to 26 August 2017 over Nigeria

Gbode, I. E., Ajayi, V. O., Adefisan, E. A., Okogbue, E. C., Cafaro, C. ORCID: https://orcid.org/0000-0001-8063-4887, Olaniyan, E. A., Ogungbenro, S. B./, Oluleye, A., Lawal, K. A., Omotosho, J. A. and Stein, T. ORCID: https://orcid.org/0000-0002-9215-5397 (2023) Verification of multiresolution model forecasts of heavy rainfall events from 23 to 26 August 2017 over Nigeria. Meteorological Applications, 30 (4). e2135. ISSN 1469-8080 doi: https://doi.org/10.1002/met.2135 Available at https://centaur.reading.ac.uk/112892/

Received: 9 May 2022 | Revised: 5 June 2023 | Accepted: 9 June 2023

DOI: 10.1002/met.2135

Meteorological Applications
Science and Technology for Weather and Climate
Open Access
RMetS

RESEARCH ARTICLE

Verification of multiresolution model forecasts of heavy rainfall events from 23 to 26 August 2017 over Nigeria

Imoleayo E. Gbode[1] | **Vincent O. Ajayi**[1] | **Elijah A. Adefisan**[1,2] | **Emmanuel C. Okogbue**[1] | **Carlo Cafaro**[3] | **Eniola A. Olaniyan**[4] | **Stephen B. Ogungbenro**[1] | **Ayodeji Oluleye**[1] | **Kamoru A. Lawal**[4,5] | **Jerome A. Omotosho**[1] | **Thorwald Stein**[6]

[1]Department of Meteorology and Climate Science, Federal University of Technology Akure, Akure, Nigeria

[2]GCRF African SWIFT Project Office, African Center Of Meteorological Applications for Development, Niamey, Niger

[3]Met Office@Reading, Department of Meteorology, Brian Hoskins Building, University of Reading, Reading, United Kingdom

[4]Deparment Weather Forecasting Services Numerical Weather Prediction, Nigerian Meteorological Agency (NiMet), Abuja, Nigeria

[5]African Climate and Development Initiative, University of Cape Town, South Africa

[6]Department of Meteorology, University of Reading, Berkshire, United Kingdom

Correspondence
Imoleayo E. Gbode, Department of Meteorology and Climate Science, Federal University of Technology Akure, Nigeria.
Email: iegbode@futa.edu.ng

Funding information
UK Research and Innovation as part of the Global Challenges Research Fund, Grant/Award Number: NE/P021077/1

Abstract

The study uses numerical weather prediction models to predict the occurrence of heavy convective rainfall associated with the passage of the African Easterly Wave (AEW) during the period 23–26 August 2017 over Nigeria. Fraction skill score (FSS) and method for object-based diagnostic evaluation (MODE) verification techniques were applied to verify how well the models predict the high-impact event and to demonstrate how these tools can support operational forecasting. Ensemble model forecasts at a convective scale from UK Met Office Unified Model (MetUM) and a one-way nested weather research and forecasting (WRF) model were compared with the integrated multisatellite retrievals for global precipitation measurement (IMERG GPM). The purpose is to examine skills of improved model resolution and ensemble in reproducing rainfall forecasts on useful scales and how the skill varies with spatial scale. WRF 2 and 6 km model forecasts show comparable skill at smaller grid scales. The skill of MetUM improves dramatically when the verification statistics are applied to the ensemble mean of the binary fields of the individual member forecast. The object-based analysis reveals a similar structure as observed, although displaced eastwards. Most improvement occurred for heavier rainfall events associated with the passage of the AEW. WRF 6 km compares reasonably well with WRF 2 km in terms of shape and structure of rainfall underscoring the ability of the model to reasonably represent convection at 6 km horizontal resolution. The ensemble members in MetUM explicitly reproduce convection at 4 km resolution but are displaced at about 166 km behind observed rainfall.

KEYWORDS

African Easterly Wave, ensemble forecasts, forecasting, fraction skill score, extreme rainfall, extremes, hazards, heavy rainfall, Met Office Unified Model, method for object-based diagnostic evaluation, precipitation, probabilities, severe weather, Weather Research and Forecasting model

1 | INTRODUCTION

The socioeconomic importance of high-impact events such as extreme rain or wind storms are phenomena resulting in loss of lives and livelihoods, damage to physical environments, and cancelation of national development gains. Nigeria is increasingly suffering from annual flooding during the rainy seasons caused by increased precipitation linked to high-impact events (Aja & Olaore, 2014). Such extreme weather events are becoming more intense while they remain difficult to forecast using numerical weather prediction models. Sub-grid scale variability, even in high-resolution models, makes it challenging to verify forecast performance. Most conventional verification scores do not characterize forecast quality at spatiotemporal scales because they only consider comparisons on a point-to-point basis (Baldwin & Kain, 2006; Casati et al., 2008; Gilleland et al., 2009; Jolliffe & Stephenson, 2012; Wilks, 2005). For instance, a good forecast with a reasonable size and structure of the feature of interest might give very poor verification when displaced slightly in space because it misses the observations and gives a false alarm, known as "double penalty" (Ebert, 2008; Gilleland et al., 2009; Rossa et al., 2008).

Predicting the occurrence of convective rainfall is, however, challenging owing to possible rapid growth of errors on small scales together with inherent errors on larger scales. This often makes higher-resolution forecasts at convection-permitting resolution less skillful, with higher probability of having a larger amount of small-scale intensity error (Bannister et al., 2017; Gilleland et al., 2009); therefore, making the traditional verification method less useful in determining the true quality of a forecast. A way of verifying forecasts using widely used statistical measures such as Pearson correlation, mean bias, mean absolute error, among others, is by synthesizing the individual statistics to create a comparative model skill score (Gbode et al., 2019). While this type of score could be more applicable for evaluating long-term averaged fields, it provides no information regarding the location and structure errors, which are of more importance to operational forecasting.

Over the past few decades, researchers have proposed numerous spatial verification methods to address the shortcomings inherent in the traditional verification methods when applied to high-resolution forecasts (Gilleland et al., 2009). These methods can basically be grouped into two major categories: filtering methods and displacement methods. The two main techniques under the filtering methods are neighborhood (or fuzzy) and scale separation (or scale decomposition) techniques while the displacement methods are the features based (or object based) and field deformation techniques (Gilleland et al., 2010; Dorninger et al., 2018). In essence, the spatial verification techniques are broadly grouped into four categories under the two major categories. Depending on the user's need and purpose, these methods give more detailed information than conventional metrics. Spatial verification methods have been used recently in Tropical Africa to assess the skill of convection-permitting ensembles (East Africa; Cafaro et al., 2021) and an operational deterministic model (West Africa; Olaniyan et al., 2022).

The current study will apply neighborhood and feature-based techniques from the filtering and displacement methods, respectively, to investigate the forecast performance based on different scales, location and structure errors, and timing errors of multiresolution model forecasts of heavy rainfall events.

The purpose of this study is to apply different spatial forecast verification methods so as to examine how forecast skill varies with spatial scale in a manner that can be comprehensively understood by users and applied to operational forecasting. The study focused on a heavy rainfall event that took place across West Africa as part of westward movement of an African Easterly Wave. The reason for selecting a single case study is to illustrate the use of this type of analysis for operational verification by forecasters. Furthermore, it is important to note that the use and implementation of convection-permitting ensembles for operational forecasting are novel in Tropical Africa (see Cafaro et al., 2021, in the references). Therefore, operational forecasters of the countries of Tropical Africa do not have any experience with the interpretation and verification of probabilistic products. Thus, we believe that the analysis of this article, even if for a single case study, can provide valuable information and enhance operational forecasting in tropical Africa. The following section describes the model configurations and verification techniques in detail. The results and discussion are presented in Section 3 and conclusion is made in Section 4.

2 | DATA AND METHODS

2.1 | Model description and setup

The individual Met Office Unified Model (MetUM) 18-member ensemble (hereafter MetUM Ens) forecast is a downscaler of the global ensemble similar to the setup used by the UK regional ensemble, that is, MOGREPS-UK (Hagelin et al., 2017), at 4 km convection-permitting limited-area covering West African domain (Figure 1a). The initial and boundary conditions for MetUM are taken from the global ensemble running at the Met Office, MOGREPS-G at ~0.28° resolution with 18 members (Bowler et al., 2008; Bowler & Mylne, 2009; Tennant & Beare, 2014). Each of the ensemble members was run for 5 days with initialization times of 0000 and 1200 UTC for the regional

ensemble covering the heavy rainfall events period, August 23–28, 2017. In this study, we only analyze the 5-day forecast initialized at 0000 UTC on August 23, 2017. A series of mesoscale convective systems (MCSs) moved over West Africa between August 19 and 26, 2017. The propagating systems along their course caused heavy rainfall, which led to flooding in numerous regions throughout the region. This flooding disaster had significant socioeconomic repercussions. Over 100,000 people were reportedly homeless, and over 4000 homes were destroyed in 12 local government areas, according to Nigeria's National Emergency Management Agency. The Niger Republic, with an estimated 200,000 displaced people, Guinea, with about 3500 affected individuals and 10 fatalities, and Ghana, with an estimated 3000 displaced persons and seven fatalities, are other nations impacted by these events (OCHA, 2017).

Another deterministic model setup was made using the WRF version 4.2 model as a common framework to perform two nested domain simulations with 18, 6, and 2 km (hereafter WRF18, WRF6, and WRF2; Figure 1a) horizontal resolutions. The outer domain covers West Africa and the innermost domain, which runs at convection-permitting scale, focuses on Nigeria (Figure 1b). When interpreting the results, it is worthy of note that scale 0 = 18 km, which is three times the grid scale for WRF6, nine times the grid scale for WRF2, and approximately four times the grid scale for MetUM Ens, that is, there is a fair degree of smoothing that has been applied using a bilinear regridding process to get the models onto a level playing field. Only WRF18 is really evaluated on its native grid and has not benefited from any additional smoothing.

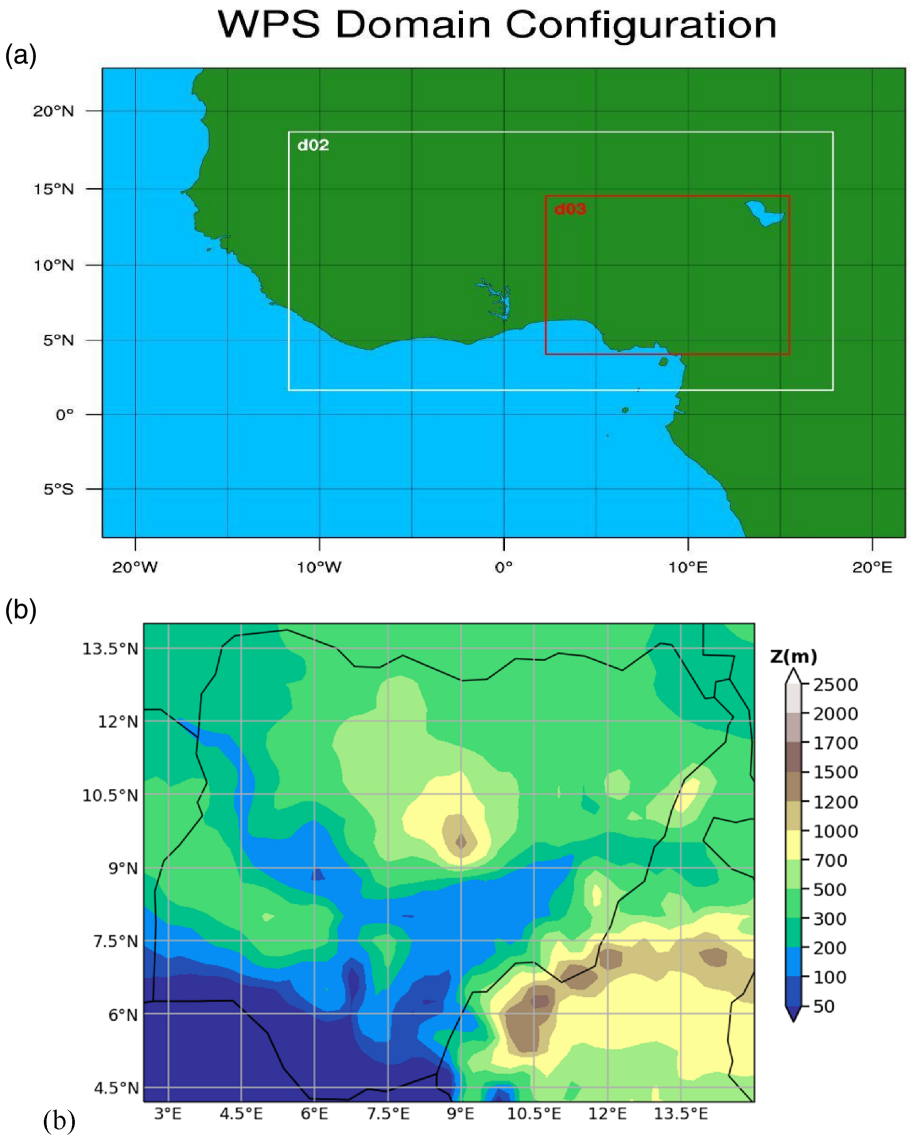

FIGURE 1 Shows maps of (a)WRF model domain with a parent domain (WRFd01) of 18 km horizontal resolution while the two inner domains (i.e., WRFd02 and WRFd03) have resolutions of 6 and 2 km, respectively, and (b) WRFd03 domain with elevation in meters.

The WRF model setup is similar to the study by Gbode et al. (2019) in terms of the model physics combination used in the model simulations. The parameterization schemes used are the Goddard (GD) WRF model microphysics (MP), the Mellor–Yamada–Janjic (MYJ) planetary boundary layer (PBL), and the Bett-Miller-Janjic (BMJ) cumulus convection (CU) parameterization schemes. This combination was found to reproduce realistic rainfall and temperature relative to gridded observations in a previous study over West Africa (Gbode et al., 2019). The GD is a six-class microphysics with graupel and modifications for ice/water saturation based on Lin et al. (1983). MYJ is a local closure scheme that predicts turbulent kinetic energy (Zhang et al., 2012) and the BMJ CU is a profile adjustment scheme that relaxes both deep and shallow profiles toward a reference profile without explicit updraft, downdraft, or cloud entrainment. However, the CU scheme was turned off in the 2 km domain to explicitly represent convection. The initial and boundary conditions used are the historical reanalysis of the Global Forecast System (GFS-ANL; https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-forcast-system-gfs) at 0.5° horizontal resolution. GFS is a weather forecast model produced by the National Centers for Environmental Prediction (NCEP). The forecast model produces a global dataset of several atmospheric and land-soil variables such as temperatures, winds, precipitation, soil moisture, and atmospheric ozone concentration at a base horizontal resolution of 18 miles (28 km) between grid points, which is used operationally to produce 16-day weather forecasts. The one-way nested WRF model forecast was initialized at 0000 UTC on 23rd and was run for 5 days till August 28, 2017. The WRF model became stable after 3 h of initialization and produced realistic forecast.

The precipitation forecasts were verified against the half-hourly final Integrated Multi-satellitE Retrievals for Global Precipitation Measurement (GPM) (IMERG; hereafter GPM) at 0.1° horizontal resolution. The estimates are derived from various precipitation-relevant satellite passive microwave (PMW) sensors comprising the GPM constellation using the 2017 version of the Goddard Profiling Algorithm (GPROF2017), then gridded, inter-calibrated to the GPM Combined Ku Radar-Radiometer Algorithm (CORRA) product, and merged into half-hourly $0.1° \times 0.1°$ (roughly 10×10 km) fields. The "Final" satellite-gauge product is produced ~3.5 months after the observation month, using both forward and backward morphing and including monthly gauge analyses, which makes it a very good alternative in a data sparse region like West Africa (Maranan et al., 2020).

2.2 | Verification techniques

The below subsections will briefly describe the two spatial verification methods used in the current study. These methods were selected from the filtering methods (i.e., neighborhood approaches) and displacement methods (i.e., feature-based (or object-based) approaches).

2.2.1 | Neighborhood techniques

The neighborhood verification techniques compare forecasts and observed values in space–time neighborhoods relative to a point of the field of interest. The approach upscales the field by averaging the values of neighbors of grid points within a certain radius of each other. This creates a smoothed field from the original field that is used to compute the summary statistics that follow closely the traditional verification statistics. The comparisons are repeated for incrementally larger neighborhoods in order to determine the scale at which a desired level of skill is attained by the forecast. Depending on the neighborhood techniques used, the approach can provide useful qualitative information, determining the optimum resolutions the forecast performs best and reducing the double-penalty problem (Ebert, 2008; Gilleland et al., 2009; Rossa et al., 2008).

One of the widely used techniques in this approach is the fractions skill score (FSS, Roberts, 2008; Roberts & Lean, 2008). FSS provides an assessment of the dependency of skill on spatial scale and intensity, thereby making it an ideal skill score for verifying high-resolution rainfall forecasts. It compares the fractional coverage of events defined by occurrences of values exceeding a given threshold in windows surrounding the observations and forecasts (Mittermaier & Roberts, 2010). The forecast and observed values are projected onto the same verification grid. User-defined thresholds (e.g., rainfall thresholds [tr] of 0.5, 1, 2, and 4 mm) are used to convert the observed (O_r) and forecast (F_r) rainfall fields into binary fields I_o and I_F (Equation 1). Any grid point where the values exceed the threshold is assigned a value of 1 and where the conditions are not met have a value of 0,

$$I_o = \begin{cases} 1, O_r \geq tr \\ 0, O_r < tr \end{cases} and\, I_F = \begin{cases} 1, F_r \geq tr \\ 0, F_r < tr \end{cases} \quad (1)$$

Also, percentile-based thresholds instead of rainfall amount thresholds can be used to generate the binary fields from forecast and observations. This removes the impact of any bias in rainfall amounts and thus allows for a good judgment of the spatial accuracy of the forecasts.

The fractions are generated to obtain the probabilities following the same nearest-neighbors method proposed by Theis et al. (2005) and Roberts (2008) as described in Equations 2 and 3, where the quantities derived are used to assess the spatial density in the binary fields. The fraction of surrounding points within a given square n that have a value of 1 was obtained for every grid point in the binary fields.

$$O(n)(i,j)=\frac{1}{n^2}\sum_{k=1}^{n}\sum_{l=1}^{n}I_o\left[i+k-1-\frac{(n-1)}{2},\ j+l-1-\frac{(n-1)}{2}\right], \tag{2}$$

$$F(n)(i,j)=\frac{1}{n^2}\sum_{k=1}^{n}\sum_{l=1}^{n}I_F\left[i+k-1-\frac{(n-1)}{2},\ j+l-1-\frac{(n-1)}{2}\right], \tag{3}$$

where $O(n)$ (i, j) and $F(n)$ (i, j) are the resultant fields of observed and forecast fractions, respectively, for a square of length n obtained from the corresponding binary field I_o and I_F. The index i ranges from 1 to N_x number of columns in the domain, and j from 1 to N_y, number of rows in the domain. The value of n can be any odd value up to $2N-1$, where N is the number of points along the longest side of the domain, and can be varied to compute fractions at different spatial scales.

For comparison, a convolution kernel for the mean filter is applied to the binary field to create squares. Thus Equation 2 becomes:

$$O(n)(ij)=\sum_{k=1}^{n}\sum_{l=1}^{n}I_o\left[i+k-1-\frac{(n-1)}{2},j+l-1-\frac{(n-1)}{2}\right]K(n)(k,l), \tag{4}$$

where $K(n)(k,l)$ is the $n\times n$ kernel for a (square) mean filter.

The mean square error (MSE) for the observed and forecast fractions from a neighborhood of length n is given by

$$MSE_{(n)}=\frac{1}{N_xN_y}\sum_{i=1}^{N_x}\sum_{j=1}^{N_y}\left[O_{(n)i,j}-F_{(n)i,j}\right]^2 \tag{5}$$

The MSE is less useful because it highly depends on the frequency of the event itself; therefore, another skill score (i.e., Fractional Skill Score [FSS]; Murphy & Epstein, 1989) based on MSE is computed relative to a low-skill reference forecast.

$$FSS_{(n)}=\frac{MSE_{(n)}-MSE_{(n)\mathrm{ref}}}{MSE_{(n)\mathrm{perfect}}-MSE_{(n)\mathrm{ref}}}=1-\frac{MSE_{(n)}}{MSE_{(n)\mathrm{ref}}}, \tag{6}$$

where $MSE_{(n)\mathrm{perfect}}=0$ is the MSE of a perfect forecast for neighborhood length n. The reference $MSE_{(n)\mathrm{ref}}$ for each neighborhood length n is given by,

$$MSE_{(n)\mathrm{ref}}=\frac{1}{N_xN_y}\left[\sum_{i=1}^{N_x}\sum_{j=1}^{N_y}O^2_{(n)i,j}+\sum_{i=1}^{N_x}\sum_{j=1}^{N_y}F^2_{(n)i,j}\right]. \tag{7}$$

The $MSE_{(n)\mathrm{ref}}$ is a measure of the largest possible MSE that can be obtained from the forecast and observed fractions.

In the current study, a bilinear regridding method was applied to both observed and modeled datasets to regrid to 18 km horizontal resolution, which is the coarsest resolution derived from the model forecast. This was done for direct grid point comparison between observation and forecast. The FSS was computed using different rainfall amount thresholds on the 3 h precipitation from 0300 UTC of 23rd to 0000 UTC of August 26, 2017. Another FSS was computed using the 95th percentile threshold instead of rainfall amount thresholds to generate binary fields from forecast and observations. This approach eliminates the effect of any bias in rainfall amounts and therefore provides a good judgment of the spatial accuracy of the forecasts.

2.2.2 | Features-based techniques

The features-based approach, also known as object-based or cell identification techniques, can be used to determine what constitutes a feature, whether spatially discontinuous features within a field should be treated as one feature or separate features, how they match features from one field to the other, and what sort of diagnostics and/or summary measures they produce (Gilleland et al., 2009). An object is identified by applying a threshold to the fields. The attributes of each of the identified objects, for example, the size, shape, and average intensity, are calculated for the forecast or observation field. These characteristics can be compared against identified features in the corresponding field in terms of their spatial proximity and summary statistics describing how well these objects compared can be calculated.

The Method for Object-Based Diagnostic Evaluation (MODE) is a feature-based technique particularly useful for application to high-resolution numerical weather prediction of high-impact weather events. MODE can determine how similar the forecast objects are to the observed

objects according to a variety of descriptive criteria including the size, shape, and average intensity (Brown et al., 2007; Bullock et al., 2016; Davis et al., 2006a, 2006b). The method uses a convolution filter and thresholding to first identify objects in gridded fields. Forecast skill at different spatial scales can be investigated by changing the values of the filter and threshold parameters. After the objects are identified, a feature-based scheme is used to merge objects within a field and match them between the forecast and the observations (Brown et al., 2007; Davis et al., 2006a, 2006b). The characteristics of the matched objects (e.g., location, area, volume, intensity, shape, etc.) are compared to examine their similarities. Information derived from these attributes is combined to give an interesting value that summarizes the goodness of the match. MODE attributes can provide details of single matched shapes (i.e., hits), single unmatched shapes (i.e., false alarms, misses), clustered objects (i.e., groups of forecast or observed objects that are merged together), and interest to specific users (e.g., distance between storms, which is relevant for aviation strategic planning), as well as details of how forecasts represent the storm/precipitation climatology, understand systematic errors, and document variability in performance in different situations.

In this study, the MODE was applied to the 18 km regridded forecasts and observation. A threshold of 5 mm rainfall amount and radius of 60 km was used to characterize the forecast and observed objects and a merge threshold of 1.25 mm was chosen. The merging technique in MODE requires that the chosen merge threshold should be less than the threshold uses to define objects, to produce larger merged objects that fully contain the originally defined objects. The total interest value of 0.7 is used to summarize the goodness of the matched objects.

3 | RESULTS AND DISCUSSION

3.1 | Spatial distribution and statistics of rainfall

First, we consider the propagation of the rainfall features in observations and the different models. Figures 2–6 show the 3-h observed and predicted rainfall distribution from GPM, MetUM, and WRF for the period 2100UTC of August 24 to 0900UTC of August 25, 2017. The period corresponds to the time when there was an east to west propagation of rainfall maximum across the domain. This convective activity is associated with the passage of an African Easterly Wave (AEW) over Nigeria (Olaniyan et al., 2022). At 21:00 UTC of August 24, 2017, GPM observed accumulated rainfall extending from the southern part of Nigeria to around latitude 10.5°N (Figure 2). Maximum values of about 40 mm hr^{-1} are observed in GPM with north-west orientation and maximum cores at approximately latitude 6°N along longitudes 10.5°E and latitude 9°N between longitudes 6 and 7.5°E. Rainfall forecasts of August 24, 2017 valid for 21:00 UTC are generally displaced behind the region of observed rainfall in MetUM and WRF. MetUM ensemble mean (MetUM-Ens) shows a wide northward spread of rainfall, mostly restricted to the eastern part of the country. Also, MetUM-Ens forecast shows an isolated feature of rainfall maximum at the location near latitude 4.5°N and longitude 9°E. Deterministic forecast from WRF 18 km (WRF18) resolution shows regions of maximum rainfall (i.e., >40 mm hr^{-1}) along longitude 13.5°E with a similar north-west orientation structure as observed in GPM. Similar structures were forecast by WRF6 and WRF2 but with less intense magnitude. Most ensemble members except MetUM11, 16, and 17 show maximum at about latitude 6°N distributed between 10.5 and 13.5°E. This region of maximum corresponds approximately to the location of maximum in the three WRF deterministic forecasts.

On August 25, 2017, at 0000 UTC, the rainfall producing system associated with AEW became well organized producing more intense rainfall (Figure 3). The observed maximum core of rainfall has propagated westward by approximately 1.5°. This westward shift deforms the north-west rainfall structure observed in the previous 3 h and the leading maximum core in the north dissociates completely from the southern core. The dissociated core forms a distinct area of isolated rainfall along longitude 5°E and lies between latitudes 9° and 10.5°N, thereby causing the feature to assume a new south-east structure with eastward tilt. Similar to the observed rainfall, the model forecasts show a more organized and intense north-west line rainfall, although with spatial differences. Also, the forecast rainfall distribution propagates westward behind the observed rainfall. MetUM-Ens shows a large rainfall spread covering the south-eastern and central region of Nigeria and maximum core along the eastern border between the country and Cameroun. The rainfall spread is given by the underlying ensemble members and reflects the positional uncertainties; therefore, this bigger rainfall footprint in the ensemble mean is due to the spatial uncertainty as reflected by the ensemble members. Forecasts from the WRF model possess similar structure as observed but are displaced to the east. Relative to WRF18, the intensity of rainfall decreases in WRF2 and WRF6 forecasts. Upon visual inspection, some individual MetUM ensemble members, especially MetUM00 (control simulation, hereafter, MetUM Ctrl) and MetUM10, are closer to observations than others.

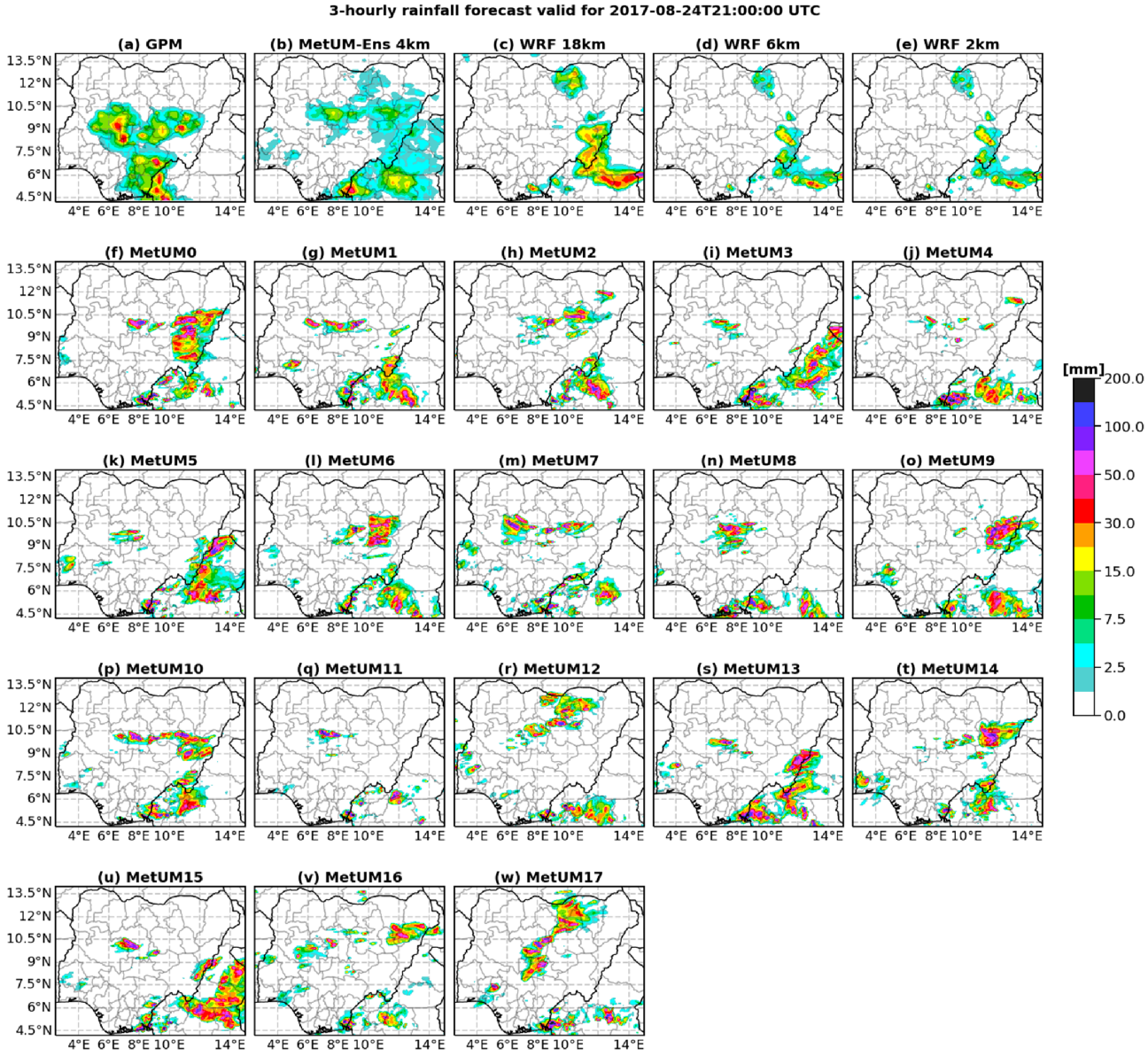

FIGURE 2 Spatial distribution of 3-h rainfall amount (mm hr^{-1}) valid for August 24, 2017 at 21:00 UTC. The first-row shows results from GPM, MetUM 18-member ensemble mean, and WRF forecast from 18, 6, and 2 km resolution deterministic forecast while rows 2–5 show the rainfall forecast from the individual MetUM ensemble member.

At 0300 UTC on August 25, 2017 (Figure 4), the observed cores of rainfall maximum propagate further to the west of the Nigerian domain. The approximate rate of westward propagation is about 1.5°/3 h. Both observed and forecast rainfall features become well-defined having a continuous line of rainfall with north-west orientation, particularly in WRF. Despite the westward movement of the rainfall core, the latitudinal position remains almost unchanged compared with the previous 3 h. Similarly, the rainfall distribution of some of the individual MetUM ensemble members (e.g., MetUM00, 01, 02, 05, 09, 10, 14, and 17) starts to develop into a well-defined structure as they transit further westwards.

At 0600 UTC of the same day (Figure 5), the observed rainfall amount has already reached the southwestern corner of the domain and values of the inland maximum core begin to reduce while the values of the coastal maximum core have increased. Though the rainfall forecast remains displaced behind the observed rainfall feature by approximately 2°, the rainfall of MetUM-Ens and the three WRF (WRF18, 6, and 2) forecasts becomes nearly identical in terms of shape and orientation. Members contributing to the more realistic feature of the MetUM-Ens include MetUM00 (MetUM Ctrl), 01, 02, 10 and 14.

As the observed system moves away from Nigeria into the Benin Republic and Atlantic Ocean to the west, an

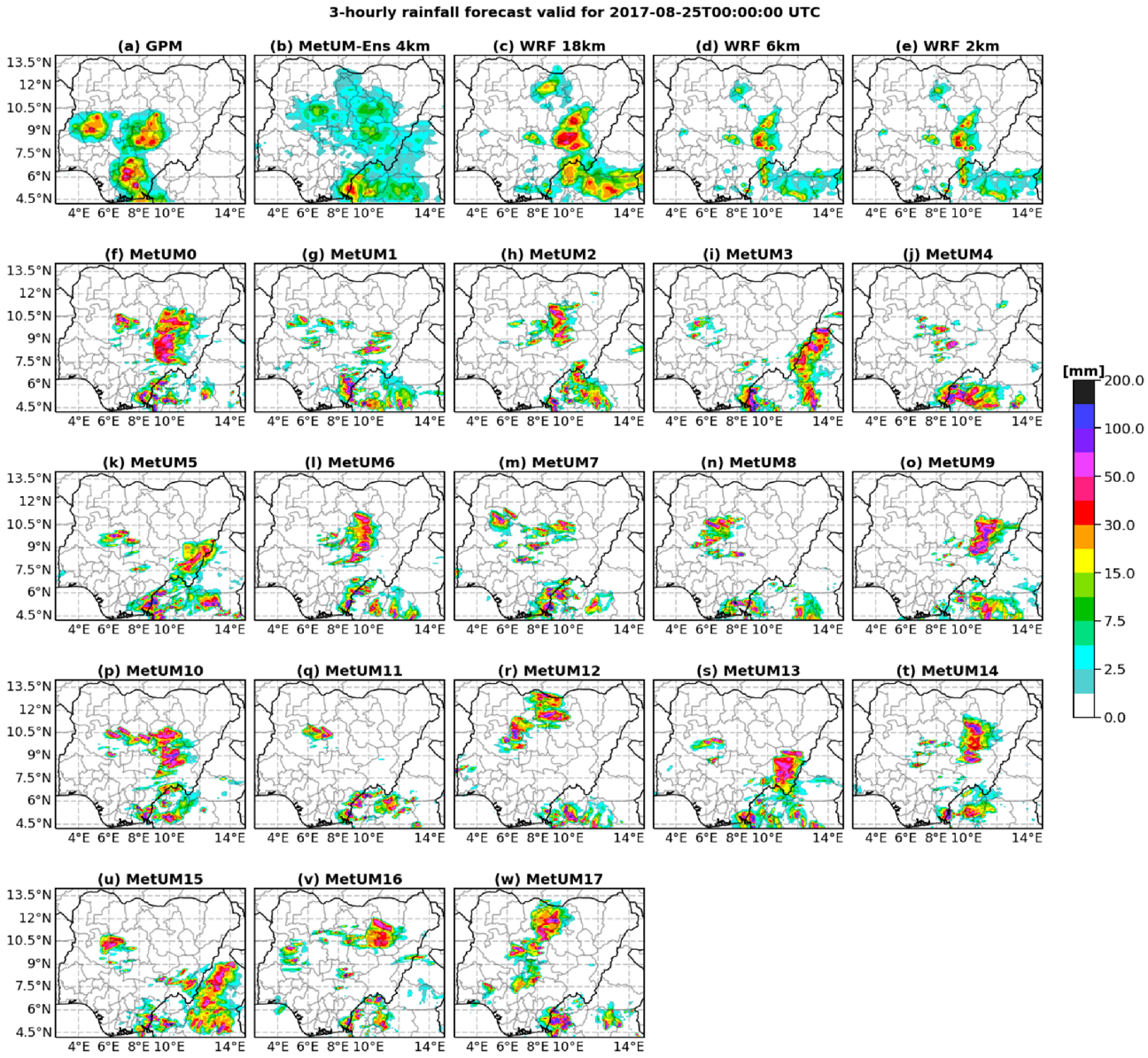

FIGURE 3 Same as Figure 2 but valid for August 25, 2017 at 00:00 UTC.

isolated rainfall-producing system not captured in the models appears to creep into the domain from the east at about 9°N (Figure 6). The forecasts start to produce reduced rainfall as it propagates further west. The similarity in WRF6 and WRF2 rainfall structure suggests that the model produced a reasonable representation of convection at 6 km horizontal resolution.

The distribution and descriptive statistics of the observed and forecast rainfall for the period August 23–25, 2017, are presented in Figure 7. The frequency of forecast rainfall of about 25 mm and below are comparable between GPM and MetUM Ctrl and WRF18, although MetUM Ctrl underestimates rainfall below and overestimates rainfall above this threshold. WRF2 and WRF6 underestimate the observed rainfall amount frequency, particularly for lower rainfall values. While MetUM Ctrl overestimates higher rainfall amounts, MetUM-Ens underestimate higher rainfall values and overestimates lower values.

3.2 | Forecast verification

3.2.1 | Fraction skill score

Figure 8 shows the FSS of different rainfall forecast thresholds for the period August 23–26, 2017. The FSS scores are derived by comparing these forecast thresholds with the observed thresholds from GPM. In general, the model forecast clearly shows that the forecast skill

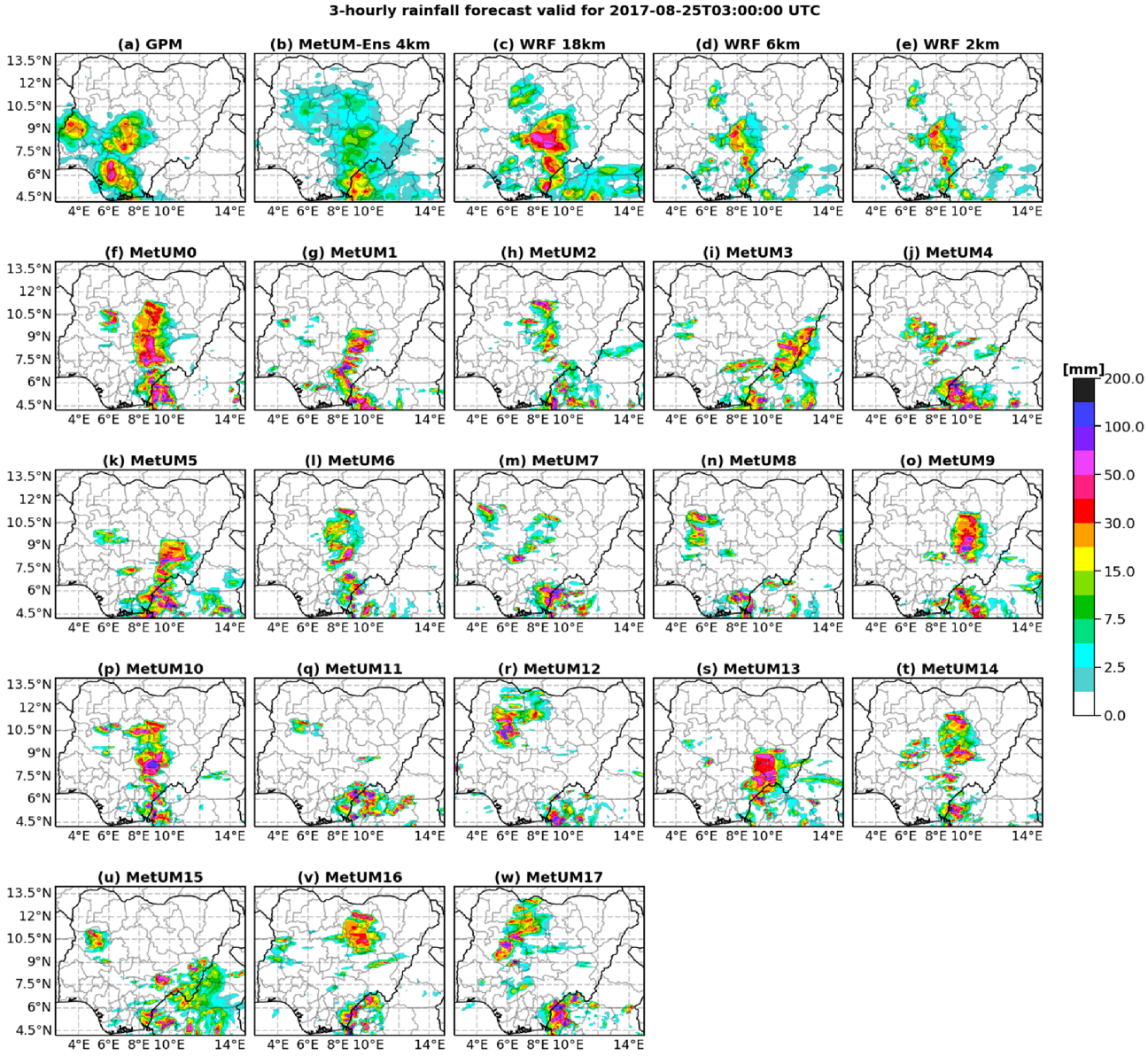

FIGURE 4 Same as Figure 2 but valid for August 25, 2017 at 03:00 UTC.

improves as the neighborhood scale increases. In WRF, the skill generally increases as the grid scale decreases, although 2 and 6 km are indistinguishable as shown in Figure 9. Also, the skill of the model forecast is higher for lower rainfall thresholds compared with higher thresholds, especially at higher neighborhood scales. For example, FSS only reaches 0.5 when the rainfall threshold is less than 10 mm in WRF and less than 12 mm in MetUM. MetUM Ens is the most skillful relative to the other models in terms of a wider range of rainfall thresholds (up to 20 mm) of FSS greater than 0.5. The area of FSS greater than 0.5 in WRF forecasts is small compared with the MetUM ensemble and control, particularly in WRF2 and WRF6 (Figure 8b, c). Although higher FSS can be seen in higher rainfall thresholds (up to 16 mm) in WRF2, this value is however attained at higher neighborhood scales (> 60). The results are consistent with Cafaro et al. (2021), that the ensemble tends to outperform the deterministic forecast, although spatial smoothing of the deterministic forecast can improve the skill.

The results from Figure 8 are affected by model biases in the rain rate distribution. As the neighborhood scale increases, the FSS should asymptote to 1 if there were no frequency bias. This can indeed be observed for the lowest rainfall rates, as Figure 7 indicates that the model frequency bias is close to 1 for those rain rates. For high rainfall rates, which are the subject of this study, frequency bias affects our interpretation of the FSS. We therefore consider FSS using a percentile threshold instead, shown in Figure 9a for the 95th percentile. The original MetUM

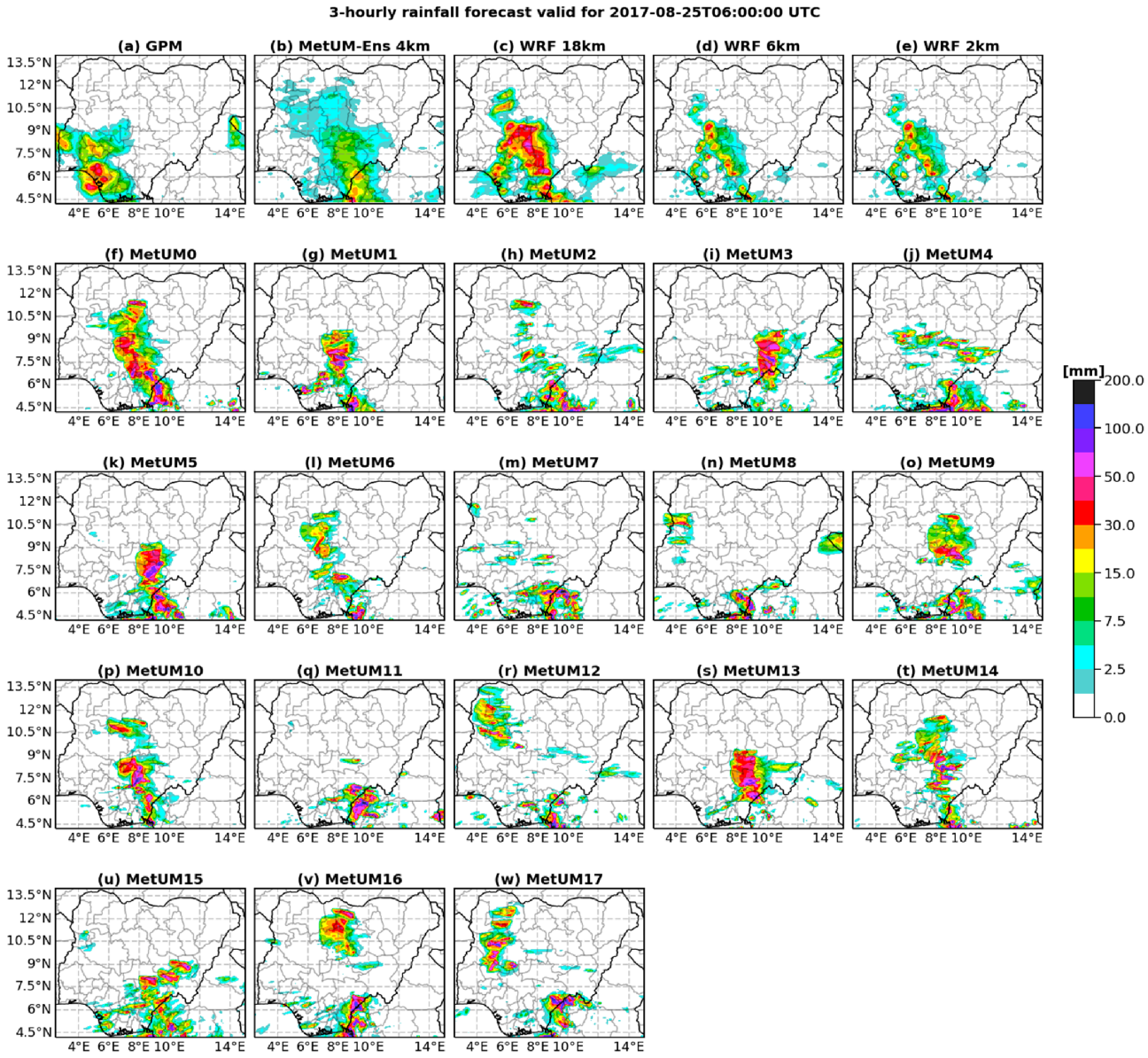

FIGURE 5 Same as Figure 2 but valid for August 25, 2017 at 06:00 UTC.

forecast (MetUM00) was considered as the control forecast (MetUM Ctrl). The ensemble mean was derived from the 17 members, and its FSS curve against spatial scales is plotted on the same chart with the ensemble FSS. Here, the 95th percentile threshold averaged over the entire domain was used to compute the scores. The deterministic FSSs of all MetUM member forecasts were also plotted for reference. The threshold used corresponds to MetUM ensemble, MetUM Ctrl, WRF18, WRF6, and WRF2. Figure 9b shows the rainfall values of the 95th percentile rainfall threshold used to generate the binary fields. FSS of MetUM ensemble shows dramatic improvement at all grid scales highlighting the benefits of applying verification score using a percentile threshold. The ensemble also performs better than most individual members. The FSS curve of each MetUM ensemble member in Figure 9 is similar to the findings of Roberts and Lean (2008) and Duc et al. (2013) where the FSS values are equal to zeros for all spatial scales less than or equal to displacement errors. Similar to the current study, their study computed ensemble FSS and the FSS of ensemble mean using all members, which is expected to show curves with zero values only when the spatial scales are smaller than the minimum displacement error of all members. According to Duc et al. (2013), this means that even when a control forecast shows an unskillful forecast via an FSS value of zero, the ensemble and ensemble mean FSS values may differ from zero, showing that the ensemble system possesses a certain level of skill in which good forecasts occur in some members different from the control.

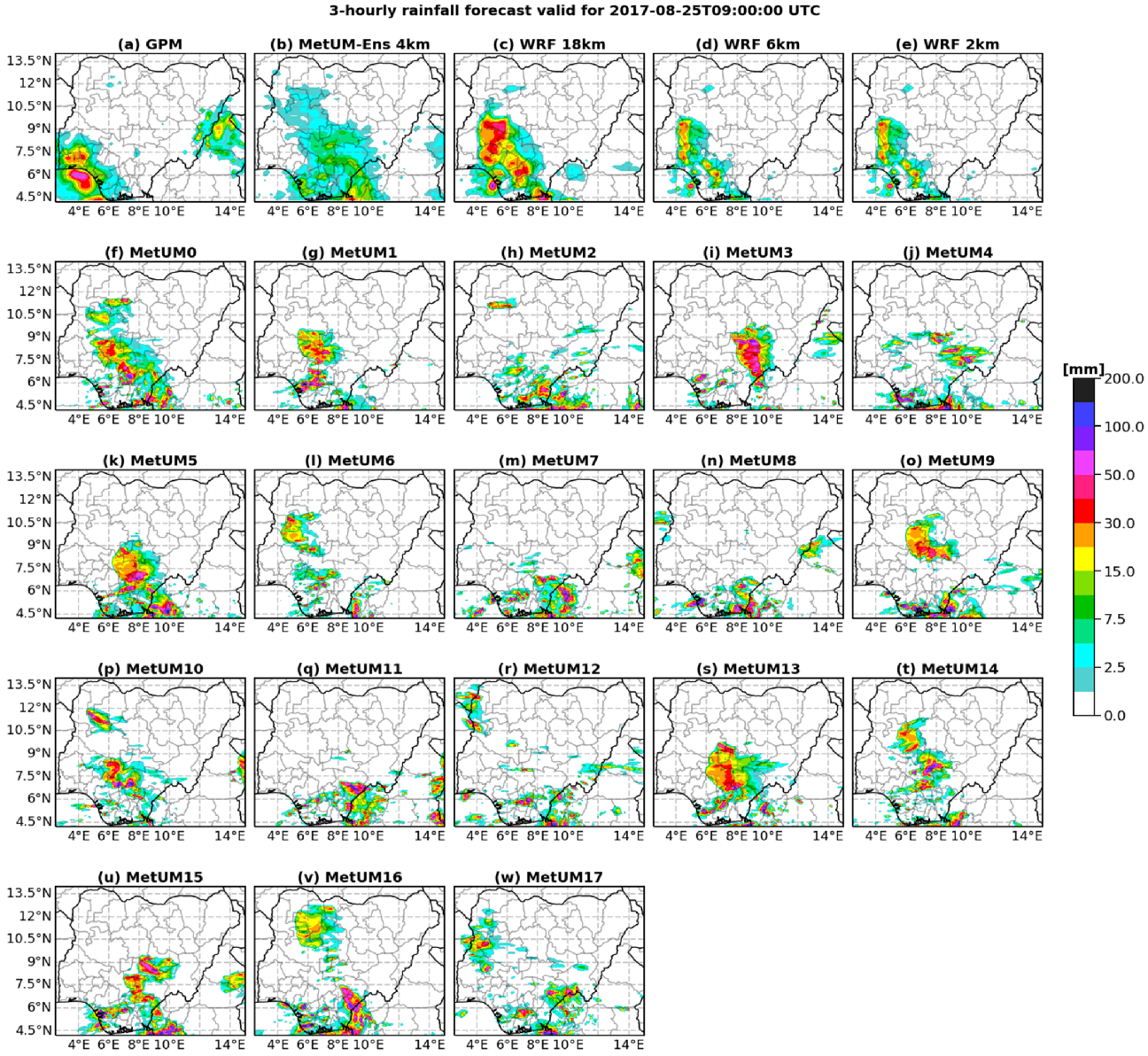

FIGURE 6 Same as Figure 2 but valid for August 25, 20177 at 09:00 UTC.

The useful scale (i.e., FSS = 0.5) at 18 km horizontal resolution is 14.5 grid points (261 km) and 18.5 grid points (333 km) for MetUM ensemble and Ctrl (Figure 9a), respectively. These values are better compared with WRF2, 6, and 18, where corresponding grid scales of 21 grid points (378 km), 22 grid points (396 km), and 24 grid points (432 km), respectively, are obtained. The actual useful scale, calculated on the original grid, might be lower than these values for all the model simulations, except for WRF18 where the original grid was used. This analysis is useful to assess the relative skill of each model compared with the others and the information about which model is the most skillful is of course necessary for the forecast users. Both WRF2 and WRF6 compare well at lower scales, but the former shows slightly higher skill. In general, MetUM Ens is more skillful than the MetUM Ctrl and WRF deterministic forecast. FSS of nearly all the individual MetUM ensemble members and WRF6 and 2 are higher than WRF18. Figure 9b shows the 3 h time series of the considered period (August 23–26, 2017). All models show comparable evolution with GPM but missed the peak time of the 95th percentile rainfall amount. During the early hours of the first day, the simulated rainfall in MetUM is comparable to GPM. The ensemble spread also shows good agreement during the first and second peaks. MetUM Ctrl overestimated the first peak but forecasts magnitude closer to the observation. Although the WRF models have their peaks close to the observed peaks, WRF18 over- and under-estimate the first and second rainfall peaks, respectively. WRF2 and 6 show a

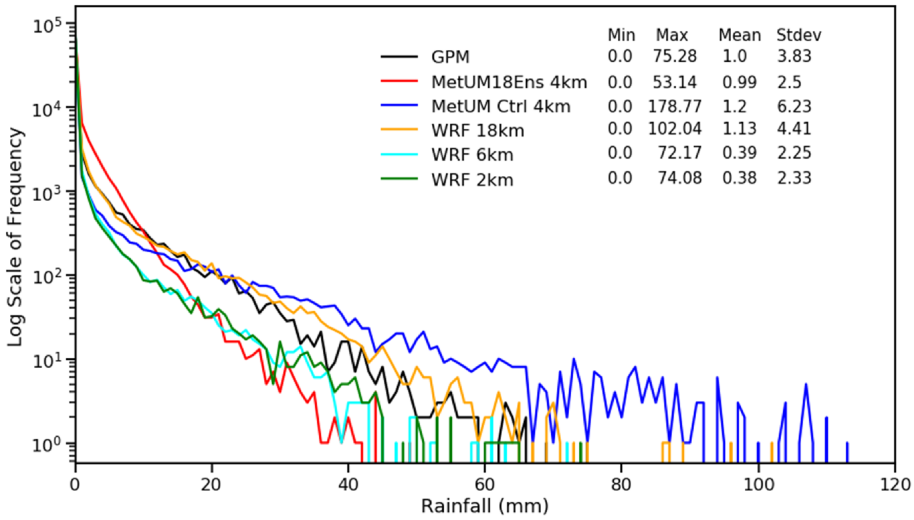

FIGURE 7 Frequency and descriptive statistics of observed and forecast 3-h rainfall amount (mm) for the period August 23–25, 2017.

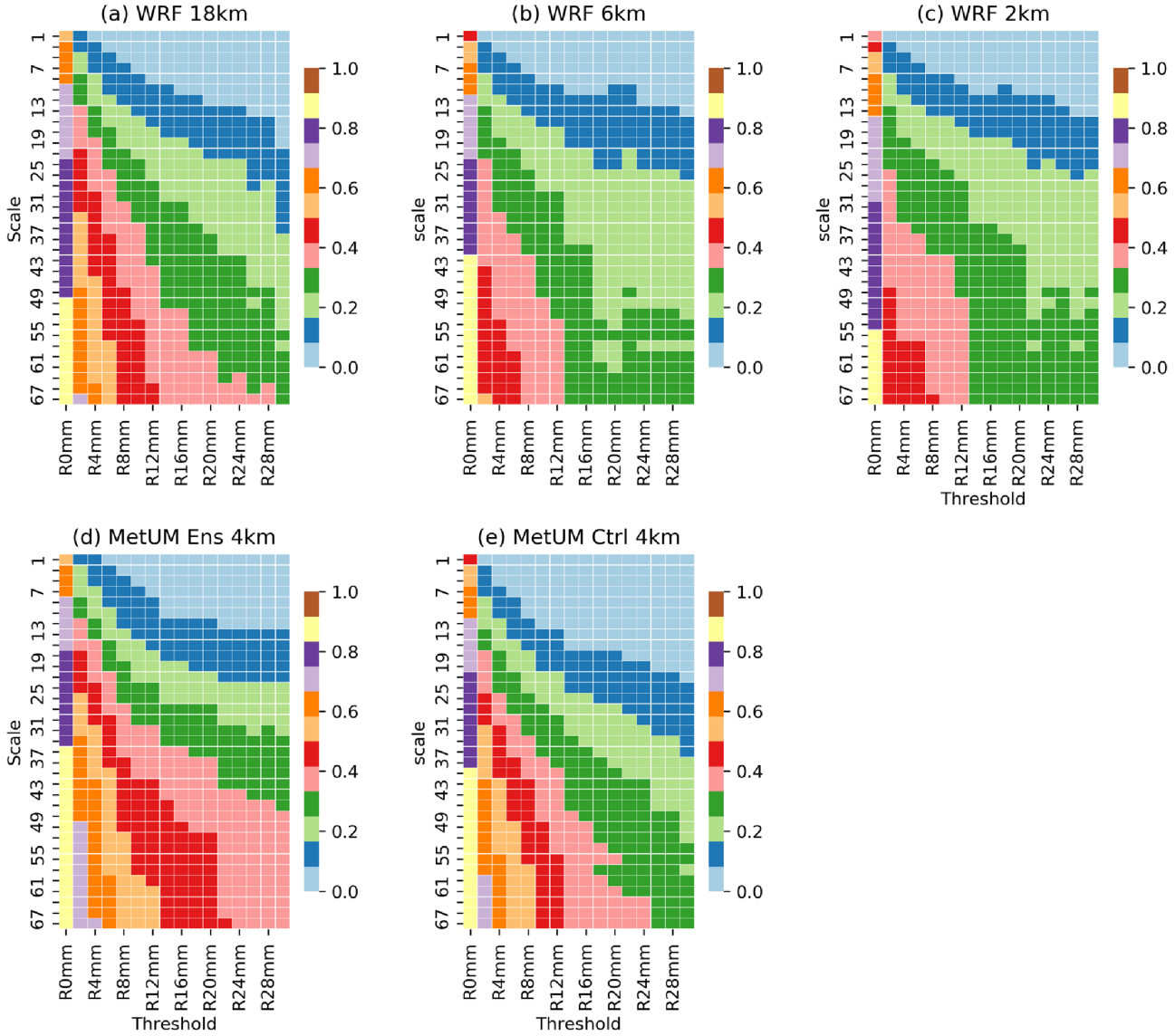

FIGURE 8 FSS of different forecast rainfall thresholds for (a) WRF18, (b) WRF6, (c) WRF2, (d) MetUM Ens (excluding the control forecast), and (e) MetUM Control run for the period August 23–26, 2017.

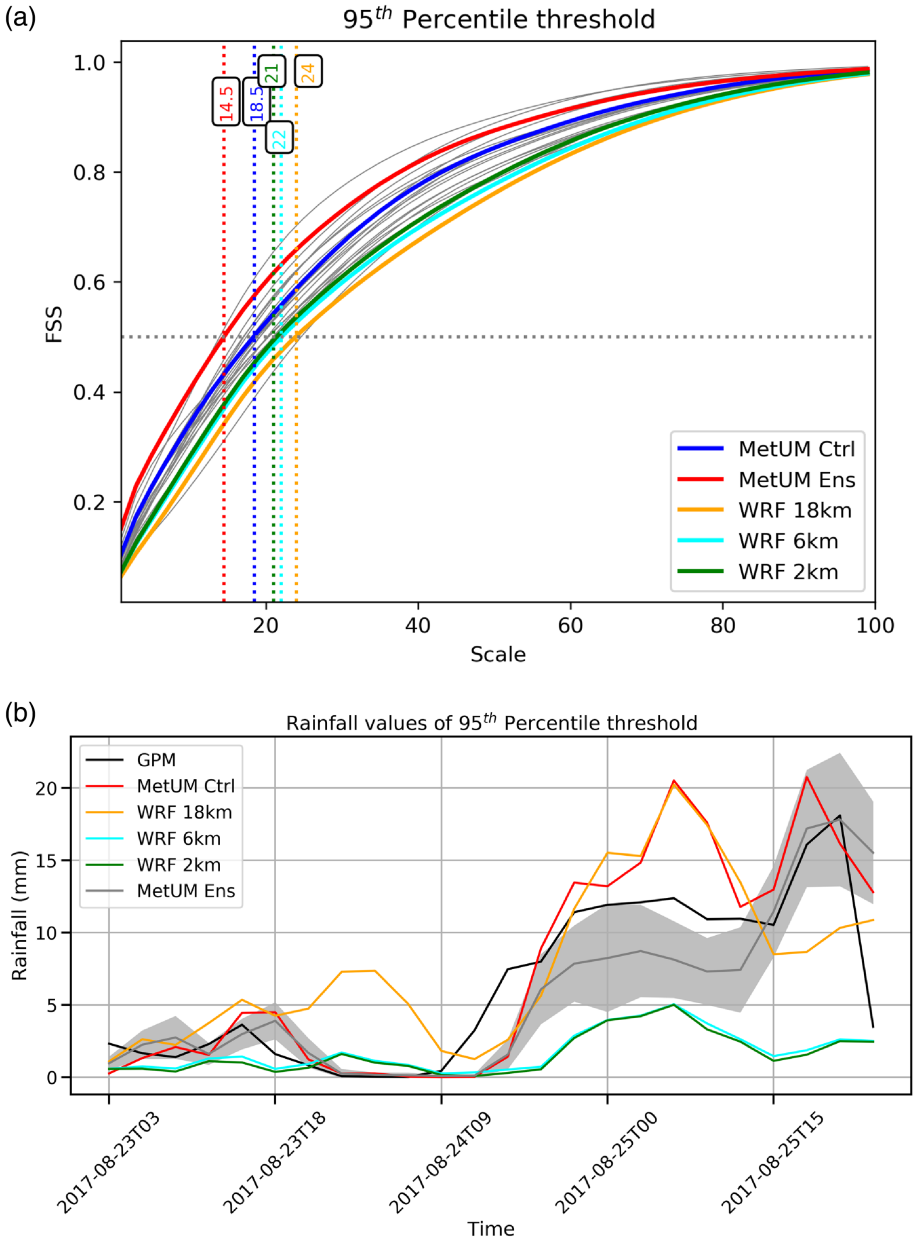

FIGURE 9 FSS of 95th percentile threshold of rainfall forecast (a) and rainfall values of 95th percentile threshold (b) for the period August 23–26, 2017. The gray lines and shades on panels a and b, respectively, represent the values of the individual MetUM ensemble member excluding the control forecast.

general underestimation of the observed 95th percentile rainfall amount.

3.2.2 | Method for object-based diagnostic evaluation

The MODE (John et al., 2022) provides an opportunity to verify the forecast in an intuitive manner such that rainfall thresholds are used to delineate forecast objects to compare with objects derived from the observed field using the same thresholds. Figure 10 shows observed and forecast features on August 25, 2017 at 00:00 UTC (compare with Figure 2). The top row shows the raw precipitation while the middle and third rows show the identified objects (shaded area) and overlay of matched forecast (blue line) and forecast objects (dashed blue line) on observed objects (red area), respectively. The matches of the objects derived from the raw precipitation fields are determined using a matched threshold greater than or equal to 1.25 mm. At this hour, the observed object is defined within the area bounded by latitudes 4°–10.5°N and longitudes 6°–9°E. Two objects were identified in the observed features. MetUM Ctrl forecast identifies four objects, but only two of these objects were matched with the observed objects. The features defined in WRF18, 6, and 2 are also found behind the observed feature. Matching the observed and forecast objects, only a fractional portion of MetUM Ctrl overlaps with the observed object. WRF18 has a broad object and lags behind the

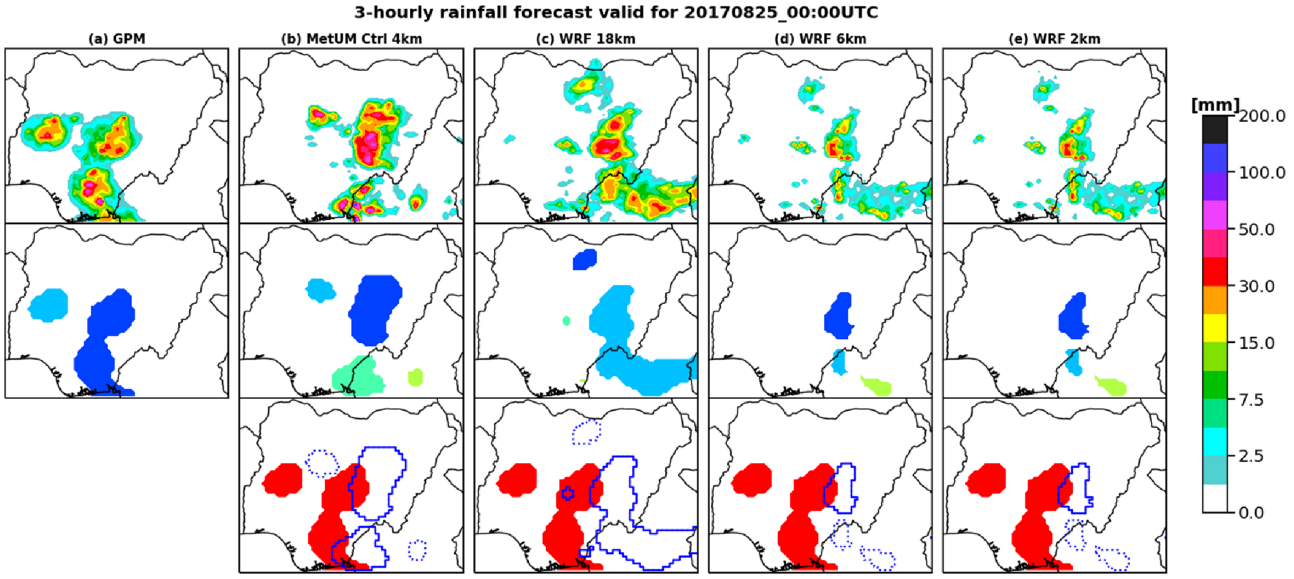

FIGURE 10 Observed and forecast features on August 25, 2017 at 00:00 UTC. The top row shows the raw rainfall amount, the middle row shows the forecast, and observed objects (shaded area) based on 5-mm rainfall threshold, and the third row presents the overlay of matched forecast (blue line) and forecast objects (dashed blue line) on observed objects (red area) determined by using 0.7 total interest threshold.

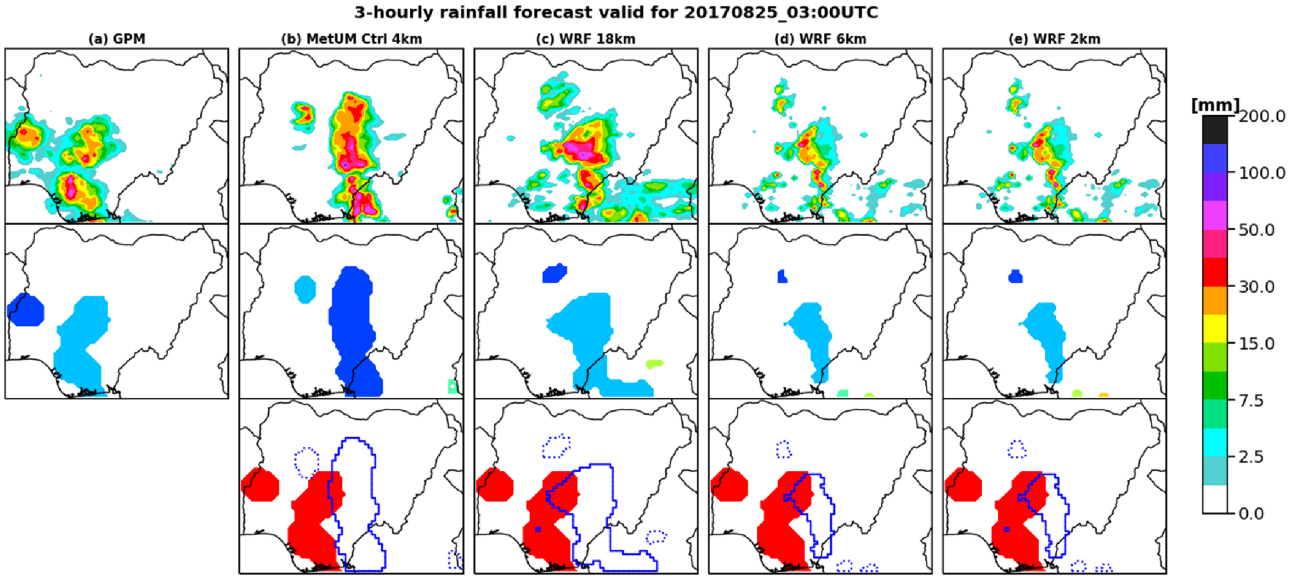

FIGURE 11 Same as Figure 10 but for August 25, 2017 at 03:00 UTC.

observed feature covering the area of the observed object around the central region of the domain. Objects of WRF6 and 2 are formed in three clusters and are almost identical with insignificant differences. Only one of these objects matches with a fractional part of the observed feature, which extends from the southern coast to the middle belt of Nigeria.

At 03:00 UTC of the same day, the observed feature propagates further to the west. The area of the MetUM Ctrl object expands and elongates north–south as it moves further west into the observed object (Figure 11). Also, the area defined by WRF18 becomes wider and matches more area of the observed feature. The area defined by WRF6 and 2 objects is narrow but matches almost the same portion of the observed objects relative to WRF18. The shape of the WRF objects extends further north of the domain. The forecast objects continue to lag behind the westward propagating observed feature. The MetUM Ctrl object is a better fit to the object though still lagging behind the observed object. An elongated

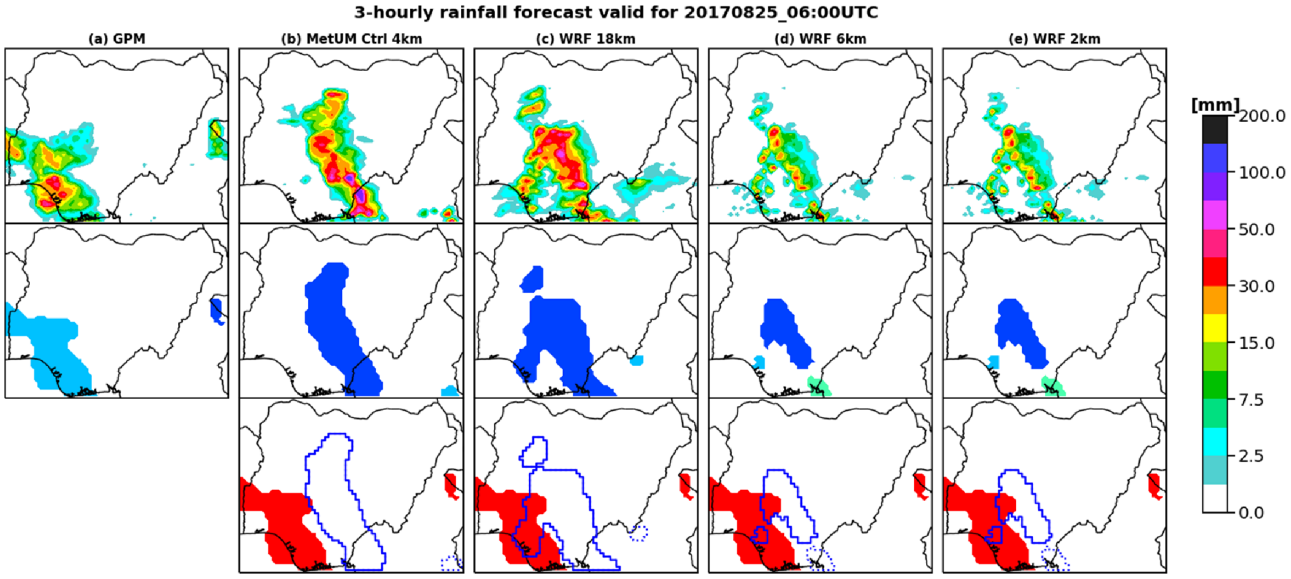

FIGURE 12 Same as Figure 10 but for August 25, 2017 at 06:00 UTC.

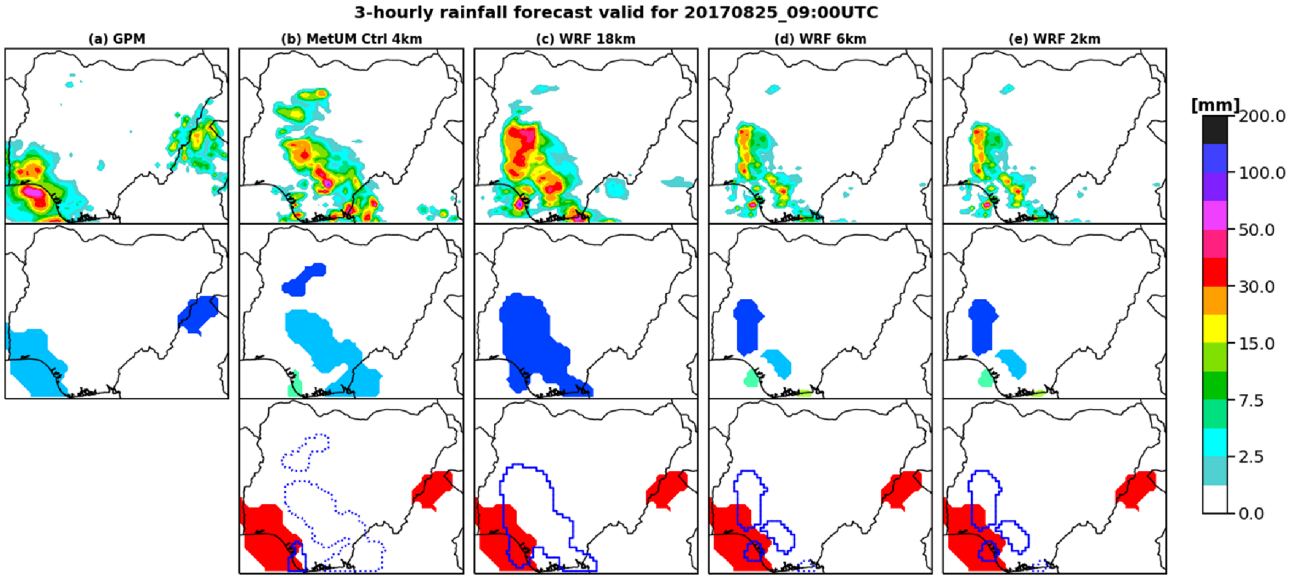

FIGURE 13 Same as Figure 10 but for August 25, 2017 at 09:00 UTC.

continuous structure is defined in WRF with a northeast orientation.

At 06:00 UTC (Figure 12), the observed rainfall activities have transited further west and resides around the southwest of Nigeria. The boundary of the MetUM Ctrl object overlays the eastern boundary of the observed object while objects defined in WRF cover larger area of the observed feature, especially in WRF18, compared with the previous hours. The rainfall activities reside around the southwest border of the country with another system entering into the country from the eastern region at 09:00UTC (Figure 13). All models were not sensitive to the encroachment of the rainfall producing system from the eastern part of the country. Only the smallest of the three objects in MetUM Ctrl matched well with the observed feature. The objects derived from the raw WRF6 and WRF2 overlap a sizable area of the observed object. Objects defined by WRF18 cover a wider area, but the largest feature delineated in MetUM Ctrl was unmatched.

The analysis of the MODE statistics, which require a careful interpretation, was omitted because of the implications of the model boundary. MODE works best when the objects are relatively small compared with the size of the domain, so that all the focus is on what goes on in the interior of the domain. Here the size of the objects relative to the domain size is comparable, which leads to

many of the objects being truncated by the domain boundaries in some way. This means that all the object attributes are affected. Therefore, none of the statistics are really usable because either the forecast or observed (or both) objects are not seen in their entirety. However, the most important outcomes from the MODE results is in terms of the positional information, which is likely to be less compromised (as here the translation appears to be more east–west than north–south), but anything else including the area ratios, unions, and complexity are all compromised in some way through the interaction/influence of the boundary.

4 | CONCLUSION

The study employs spatial forecast verification methods, FSS and MODE, to verify the skills of two NWP models in predicting the occurrence of heavy convective rainfall associated with the passage of the AEW during the period of August 23–26, 2017 over Nigeria. The purpose is to verify how well the models predict a high-impact event using two different spatial forecast verification methods so as to examine how forecast skill varies with spatial scale in a manner that can be comprehensively understood by users and applied to operational forecasting. Rainfall from ensemble model forecast at convection-permitting scale from MetUM and one-way nested WRF model were compared with the GPM IMERG. Results show that the forecast of the core of rainfall amount is displaced by about 1–1.5° behind the observed GPM rainfall in all the model forecasts. During the considered time frame, most of the individual MetUM members and WRF forecast produces convective rainfall with spatial and temporal biases. This underscores the need to further improve the model tuning and proper initialization of MetUM in order to improve the forecast. The MetUM ensemble could, however, improves the skill as was seen in the FSS statistics derived from the ensemble mean of the binary fields, as long as percentile thresholds are considered. WRF model forecasts clearly demonstrate skills in terms of forecasting reasonable shape and structure of rainfall distribution at 18, 6, and 2 km horizontal resolution, although with location displacement. This highlights the benefit of proper model setup as demonstrated in Gbode et al. (2019). Rainfall forecasts from WRF6 and WRF2 have almost the same shape and structure, suggesting that the model produced a reasonable representation of convection at 6 km horizontal resolution. The useful scales (i.e., FSS = 0.5) at 18 km horizontal resolution are 14.5 grid points (261 km) for MetUM Ens, 18.5 grid points (333 km) for MetUM Ctrl, 21 grid points (378 km) for WRF2, 22 grid points (396 km) for WRF6, and 24 grid points (432 km) for WRF18.

WRF 2 and 6 km model forecasts show comparable skills at lower neighborhood grid scales. The notable improvement seen in MetUM when the FSS verification statistics are applied to the ensemble mean of the binary fields of the individual member forecast highlights the benefit of the model ensembles. Also, the object-based analysis reveals a similar structure as observed, although displaced eastwards. The individual members of MetUM have difficulties in reproducing convective rainfall at appropriate time and location, which is also the case for WRF.

The results from this study highlight the potential of convection-permitting ensemble models to produce quality forecasts with respect to deterministic coarser resolution forecasts. Also, it shows that the use of verification techniques can be beneficial to inform operational weather forecasts and decision-making processes. It is, however, expected that numerical weather predictions of high-impact weather will continue to improve over time as quality and quantity of observation data, computational power, and knowledge of atmospheric processes improve in tropical African countries. While FSS and MODE seemingly contradict each other, they supply complementary information, with the object-based approach of MODE providing further details on displacement of individual storms. Therefore, while limited to a single case study, this study shows also the importance of applying several verification scores in order to have a more complete picture of the forecast quality. This will hopefully encourage more and deeper analysis in the future to issue better weather warnings of high-impact weather.

AUTHOR CONTRIBUTIONS

Imoleayo E. Gbode: Conceptualization (equal); data curation (equal); methodology (equal); visualization (equal); writing – original draft (equal). **Vincent O. Ajayi:** Conceptualization (equal); supervision (equal). **Elijah A. Adefisan:** Supervision (equal). **Emmanuel C. Okogbue:** Supervision (equal). **Carlo Cafaro:** Methodology (equal). **Eniola Olaniyan:** Methodology (equal). **Stephen B. Ogungbenro:** Conceptualization (equal). **Ayodeji Oluleye:** Supervision (equal). **kamoru Lawal:** Supervision (equal). **Jerome A. Omotosho:** Supervision (equal). **Thorwald Stein:** Methodology (equal); supervision (equal).

ACKNOWLEDGEMENTS

This work was supported with grants from the U.K. Research and Innovation as part of the Global Challenges Research Fund under the African Science for Weather Information and Forecasting Techniques (GCRF African-SWIFT) (work packages R2) program—grant

number NE/P021077/1. Many thanks to Dr. Jimy Dudhia and NCAR's MMM Laboratory for supporting the research work with high-performance computing resources used for the WRF simulations from Cheyenne (doi: https://doi.org/10.5065/D6RX99HX) provided by NCAR's Computational and Information Systems Laboratory, sponsored by the National Science Foundation. Also, thanks to the anonymous reviewers for the insightful comments and suggestions that helped to improve the quality of the paper. Additional computing resources and MetUM data were provided by the National Centre for Atmospheric Science JASMIN under the African-SWIFT account.

DATA AVAILABILITY STATEMENT

Model data used for this study are available on the Met Office Managed Archive Storage System (MASS) with the following path: moose:/devfc/u-be957 for MetUM. More information on how to get access to the data can be found at https://www.ceda.ac.uk/blog/access-to-the-met-office-mass-archive-on-jasmin-goes-live/. The WRF model data can be accessed on Zenodo (DOI: https://doi.org/10.5281/zenodo.8098471). The GPM-IMERG data were provided by the NASA Goddard Space Flight Center's Precipitation Measurement Missions Science Team and Precipitation Processing System, which develop and compute the GPM IMERG as a contribution to GPM, including data archiving at the NASA GES DISC.

ORCID

Imoleayo E. Gbode https://orcid.org/0000-0001-7519-512X
Eniola A. Olaniyan https://orcid.org/0000-0003-0743-9860
Thorwald Stein https://orcid.org/0000-0002-9215-5397

REFERENCES

Aja, G.N. & Olaore, A.Y. (2014) The impact of flooding on the social determinants of health in Nigeria: a case for north-south institutional collaboration to address climate issues. *Developing Country Studies*, 4(22), 6–12.

Baldwin, M.E. & Kain, J.S. (2006) Sensitivity of several performance measures to displacement error, bias, and event frequency. *Weather and Forecasting*, 21, 636–648.

Bannister, R.N., Migliorini, S., Rudd, A.C. & Baker, L.H. (2017) Methods of investigating forecast error sensitivity to ensemble size in a limited-area convection-permitting ensemble. *Geoscientific Model Development Discussion*, 1–38. Available from: https://doi.org/10.5194/gmd–2017–260

Bowler, N.E., Arribas, A., Mylne, K.R., Robertson, K.B. & Beare, S.E. (2008) The MOGREPS short-range ensemble prediction sys-tem. *Quarterly Journal of the Royal Meteorological Society*, 134, 703–722. Available from: https://doi.org/10.1002/qj.234

Bowler, N.E. & Mylne, K.R. (2009) Ensemble transform Kalman filter perturbations for a regional ensemble prediction model. *Quarterly Journal of the Royal Meteorological Society*, 135, 757–766. Available from: https://doi.org/10.1002/qj.404

Brown, B.G., Bullock, R., Gotway, J.H., Ahijevych, D., Davis, C., Gilleland, E. et al. (2007) Application of the MODE object-based verification tool for the evaluation of model precipitation fields. In 22nd Conf. on Weather Analysis and Forecasting/18th Conf. on Numerical Weather Prediction.

Bullock, R.G., Brown, B.G. & Fowler, T.L. (2016) Method for object-based diagnostic evaluation. Tech Rep.

Cafaro, C., Woodhams, B.J., Stein, T.H., Birch, C.E., Webster, S., Bain, C.L. et al. (2021) Do convection-permitting ensembles lead to more skillful short-range probabilistic rainfall forecasts over tropical East Africa? *Weather and Forecasting*, 36(2), 697–716.

Casati, B., Wilson, L.J., Stephenson, D.B., Nurmi, P., Ghelli, A., Pocernich, M. et al. (2008) Forecast verification: current status and future directions. *Meteorological Applications*, 15(1), 3–18.

Davis, C., Brown, B. & Bullock, R. (2006a) Object-based verification of precipitation forecasts. Part I: methods and application to mesoscale rain areas. *Monthly Weather Review*, 134, 1772–1784.

Davis, C.A., Brown, B.G. & Bullock, R.G. (2006b) Object-based verification of precipitation forecasts, part II: application to convective rain systems. *Monthly Weather Review*, 134, 1785–1795.

Dorninger, M., Gilleland, E., Casati, B., Mittermaier, M.P., Ebert, E.E., Brown, B.G. et al. (2018) The setup of the MesoVICT project. *Bulletin of the American Meteorological Society*, 99(9), 1887–1906. Available from: https://doi.org/10.1175/BAMS-D-17-0164.1

Duc, L., Saito, K. & Seko, H. (2013) Spatial-temporal fractions verification for high-resolution ensemble forecasts. *Tellus A: Dynamic Meteorology and Oceanography*, 65(1), 18171.

Ebert, E.E. (2008) Fuzzy verification of high-resolution gridded forecasts: a review and proposed framework. *Meteorology Applications*, 15, 51–64.

Gbode, I.E., Dudhia, J., Ogunjobi, K.O. & Ajayi, V.O. (2019) Sensitivity of different physics schemes in the WRF model during a west African monsoon regime. *Theoretical and Applied Climatology*, 136(1–2), 733–751. Available from: https://doi.org/10.1007/s00704-018-2538-x

Gilleland, E., Ahijevych, D., Brown, B.G., Casati, B. & Ebert, E.E. (2009) Intercomparison of spatial forecast verification methods. *Weather and Forecasting*, 24(5), 1416–1430.

Gilleland, E., Ahijevych, D. A., Brown, B. G., & Ebert, E. E. (2010). Verifying forecasts spatially. *Bulletin of the American Meteorological Society*, 91(10), 1365–1376. Available from: https://doi.org/10.1175/2010bams2819.1

Hagelin, S., Son, J., Swinbank, R., McCabe, A., Roberts, N. & Tennant, W. (2017) The met Office convective-scale ensemble, MOGREPS-UK. *Quarterly Journal of the Royal Meteorological Society*, 143(708), 2846–2861.

John hg, R., Hsoh-u, L., Jprestop, N., TaraJensen, G.M.C., Keith Searight, T.F., Bikegeek, T.B. et al. (2022) dtcenter/MET: MET-10.1.2 (v10.1.2). Zenodo. https://doi.org/10.5281/zenodo.6555084

Jolliffe, I.T. & Stephenson, D.B. (2012) *Forecast verification. A Practitioner's guide in atmospheric science*. Hoboken, NJ: John Wiley and Sons, p. 240.

Lin, Y.L., Farley, R.D. & Orville, H.D. (1983) Bulk parameterization of the snow field in a cloud model. *Journal of Applied Meteorology and Climatology*, 22(6), 1065–1092.

Maranan, M., Fink, A.H., Knippertz, P., Amekudzi, L.K., Atiah, W.A. & Stengel, M. (2020) A process-based validation of GPM IMERG and its sources using a mesoscale rain gauge network in the west African forest zone. *Journal of Hydrometeorology*, 21(4), 729–749.

Mittermaier, M. & Roberts, N. (2010) Intercomparison of spatial forecast verification methods: identifying skillful spatial scales using the fractions skill score. *Weather and Forecasting*, 25(1), 343–354.

Murphy, A.H. & Epstein, E.S. (1989) Skill scores and correla- tion coefficients in model verification. *Monthly Weather Review*, 117, 572–581.

Olaniyan, E.A., Cafaro, C., Ogungbenro, S.B., Gbode, I.E., Ajayi, V.O., Oluleye, A. et al. (2022) Performance evaluation of a high-resolution regional model over West Africa for operational use: a case study of august 2017. *Meteorological Applications*, 29(4), e2080. Available from: https://doi.org/10.1002/met.2080

Roberts, N.M. (2008) Assessing the spatial and temporal variation in the skill of precipitation forecasts from an NWP model. *Meteorological Applications*, 15(1), 163–169. https://doi.org/10.1002/met.57

Roberts, N.M. & Lean, H.W. (2008) Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Monthly Weather Review*, 136, 78–97. Available from: https://doi.org/10.1175/2007MWR2123.1

Rossa, A., Nurmi, P. & Ebert, E. (2008) Overview of methods for the verification of quantitative precipitation forecasts. In: *Precipitation: advances in measurement, estimation and prediction*. Berlin, Heidelberg: Springer, pp. 419–452. Available from: https://doi.org/10.1007/978-3-540-77655-0_16

Tennant, W. & Beare, S. (2014) New schemes to perturb sea-surface temperature and soil moisture content in MOGREPS. *Quarterly Journal of the Royal Meteorological Society*, 140, 1150–1160. Available from: https://doi.org/10.1002/qj.2202

Theis, S.E., Hense, A. & Damrath, U. (2005) Probabilistic pre cipitation forecasts from a deterministic model: a pragmatic approach. *Meteorological Applications*, 12, 257–268.

United Nations Office for the Coordination of Humanitarian Affairs (OCHA). (2017) West and Central Africa: 2017 flood impact. https://reliefweb.int/report/niger/west-and-central-africa-2017-flood-impact-18-oct-2017/

Wilks, D.S. (2005) *Statistical methods in the atmospheric sciences. An introduction*, 2nd edition. Burlington, MA: Academic Press, p. 648.

Zhang, C., Wang, Y., Lauer, A. & Hamilton, K. (2012) Configuration and evaluation of the WRF model for the study of Hawaiian regional climate. *Monthly Weather Review*, 140(10), 3259–3277.

How to cite this article: Gbode, I. E., Ajayi, V. O., Adefisan, E. A., Okogbue, E. C., Cafaro, C., Olaniyan, E. A., Ogungbenro, S. B., Oluleye, A., Lawal, K. A., Omotosho, J. A., & Stein, T. (2023). Verification of multiresolution model forecasts of heavy rainfall events from 23 to 26 August 2017 over Nigeria. *Meteorological Applications*, *30*(4), e2135. https://doi.org/10.1002/met.2135