Meteorology and Atmospheric Physics (2021) 133:479–493
https://doi.org/10.1007/s00703-020-00761-2

ORIGINAL PAPER

Investigating the efficacy of a new symmetric index of agreement for evaluating WRF simulated summer monsoon rainfall over northeast India

Aniket Chakravorty[1] · Rekha Bharali Gogoi[1] · Shyam Sundar Kundu[1] · P. L. N. Raju[1]

Received: 16 October 2019 / Accepted: 6 October 2020 / Published online: 19 October 2020

Abstract
The efficacy of the standard performance metrics [mean bias (Bias), root mean square deviation (RMSD), and correlation coefficient (CC)] compared to a new symmetric index of agreement (λ) for the evaluation of numerical weather prediction models is investigated in this study. It evaluates the weather research and forecasting (WRF) model with the global precipitation measurement's (GPM) Integrated Multi-satellitE Retrievals for GPM (IMERG) and station rainfall, as the reference datasets. This study uses three IMERG products, namely: GPM infrared-microwave merged gauge corrected (GMS), GPM microwave-calibrated infrared (GIR), and GPM inter-calibrated microwave (GMW) rainfall. The analysis showed that WRF rainfall when compared to different reference datasets is producing similar RMSD values but significantly different Bias values. This behavior is because of the inverse relationship between Bias and standard deviation of residual (σ_R). It is so because RMSD is a function of both. However, λ is able to appropriately represent the distinct performances of WRF. The regions with contradictory behavior of RMSD and CC are also appropriately represented in λ. The evaluation using λ showed that WRF is comparable to GMS and GIR, except for GMW. The performance of WRF was not found to be very promising when compared to station rainfall, which is attributed to WRFs representation efficiency and the effect of topography. However, a comparison of IMERG products with station rainfall showed that GMS was the most agreeable followed by GIR and GMW. The study also showed that the efficacy of λ is related to its non-linear relationship with Bias and CC.

1 Introduction

The field of numerical weather prediction (NWP) has seen a multi-dimensional growth since its advent (Abbe 1901; Bauer et al. 2015; Carpenter 1979; Dudhia 2014; Richardson 1922). NWP models have become quite popular because of their ability to generate continuous and reliable data with sufficient forecast lead time (Kalnay et al. 1990; Lynch 2008; Shuman 1989). However, weather phenomena are chaotic and non-linear, and hence, numerical modelling of these phenomena requires more than a few assumptions and simplifications. Incorporating these assumptions into the modelling structure is one of the reasons for the model error (Bjerknes 1910; Orrell et al. 2001; Tolstykh and Frolov 2005). This realisation has encouraged model evaluation studies of weather events with in-situ measurements, satellite-derived products or both (Giannaros et al. 2013; Kotroni et al. 2011; Rife and Davis 2005). A significant portion of the weather forecasting activities worldwide focus on the simulation of rainfall because of the critical role it plays in the daily routine of people. Thus, evaluation studies of simulated rainfall are of particular importance (Ebert et al. 2007; Ghile and Schulze 2010; Shrestha et al. 2013; Vasić et al. 2007).

An evaluation study estimates the degree of closeness or agreement of the simulated state to the observed state of the system. The notion of closeness or "agreement" is used to determine the difference between a model and a reference product. The metrics used for measuring the degree of agreement are categorized based on the type of data. The data can be classified into either categorical data or continuous data. Rainfall can be used as both categorical and continuous data, representing different characteristics (Bharti and Singh 2015). The frequency of occurrence of a

Responsible Editor: Emilia Kyung Jin.

✉ Aniket Chakravorty
chakravorty.aniket@gmail.com

[1] Space and Atmospheric Sciences Division, Department of Space, North Eastern Space Applications Centre, Government of India, Umiam, Meghalaya 793103, India

rainfall event, defined as rainfall amount below or beyond a threshold, represents a categorical data. However, rainfall amount over a period of time refers to a continuous data. The validation of categorical data is performed by constructing a contingency table and estimating validation metrics like, probability of detection (POD), false alarm ratio (FAR), and equitable threat score (ETS) from it. The performance metrics such as correlation coefficient (CC), coefficient of determination (R^2), root mean square deviation (RMSD), and mean bias (Bias) are used for the validation of continuous data (Wilks 2011). This study focuses on the validation measures of continuous data. These validation measures are extensively used in studies involving the analysis of rainfall amount (Adnan et al. 2020; He et al. 2017; Mitra et al. 2018; Paredes-Trejo et al. 2017; Wang et al. 2017) and thus it is pertinent to understand the dependencies between them and their inherent limitations. One of the significant limitations is that, they have to be considered simultaneously to make sense of the forecast error. For example, a simulation can have a strong correlation (CC) and also a substantial deviation from observation (RMSD and Bias). Thus, considering just one of them would lead to ambiguity in the analysis. An innovative method for analysing multiple performance metrics was suggested by Taylor (2001). However, it does not consolidate the individual measures into a single parameter. This lack of a single holistic performance measure comes at the price of simplicity. Simplicity is essential, because weather forecasting services are vital inputs to various crucial policy decisions, and a single metric based performance measure can help policymakers to get a concise summary of model skill.

The importance of a holistic metric that incorporates agreement in the temporal pattern along with information on the magnitude of deviations can be judged by the number of dedicated indices of agreement available in hydrology and remote sensing (Mielke 1984, 1991; Nash and Sutcliffe 1970; Willmott 1981; Watterson 1996; Willmott et al. 2012). However, these indices do not account for symmetry as an intrinsic property. Symmetry is a relevant property in the case of inter-comparison studies, where it is not possible to identify a priori which of the two datasets are better. Considering this, the index of agreement proposed by Mielke (1984) was modified by Duveiller et al. (2016) to incorporate the property of symmetry into it. However, the applicability of this new symmetric index of agreement (λ) as a holistic performance metric for evaluating simulated rainfall need to be further scrutinized. The present study investigates this by comparing λ against CC, RMSD, and Bias, for comparing NWP simulated rainfall with satellite rainfall. This inter-comparison between NWP and satellite rainfall is important, because more often than not it is required to estimate the performance of model rainfall in areas with a paucity of in-situ rainfall data. This is only possible by comparing it with satellite rain. The property of symmetry becomes important here as it is known that both NWP and satellite rainfall has their own uncertainties which varies from region to region.

The investigation is conducted over the complex terrain of north eastern region (NER) of India between the NWP simulated rainfall, satellite rainfall and in-situ rainfall. From the many available NWP models the weather research and forecasting (WRF) model (Skamarock et al. 2008) is selected for its reliable performance over diverse regions (Jiménez and Dudhia 2013; Kryza et al. 2013; Powers 2007; Powers et al. 2017; Paul et al. 2018). The ability of WRF has been extensively investigated, over India, and found to be appreciably reliable in various scenarios (Chawla et al. 2018; Kumar et al. 2014; Mohan and Bhati 2011; Madala et al. 2014; Srinivas et al. 2013). However, limited research has been done to examine its validity and behaviour over topographically undulating NER with satellite and in-situ rainfall estimates. Based on the above discussion, the present work is structured in a way that answers two questions: (1) What is the applicability of the new symmetric index of agreement in evaluating the forecasted rainfall amount? (2) What is the skill of WRF in simulating rainfall in northeast India?

The manuscript has been structured in the following manner. The description of the model, satellite, and in-situ products is provided in Sect. 2. The methodology followed in this study has been discussed in Sect. 3. The results of this study are discussed in Sect. 4. The conclusions from this study are discussed in Sect. 5.

2 Data description

The NWP model, satellite rainfall datasets, and in-situ station characteristics (location, altitude, and climate) are being discussed in this section. The investigation is performed for the 2017 monsoon period from June to September. The 2017 monsoon rainfall is intriguing because of its heterogeneous distribution over NER. The seasonal rainfall was deficient in three out of four meteorological sub-divisions and excess in the remaining sub-division, compared to the long period average. The meteorological sub-division of Nagaland-Manipur-Mizoram-Tripura (NMMT) was the only sub-division in India with an excess monsoonal rainfall (IMD, 2017 Southwest Monsoon End of Season Report). In contrast, a homogeneous rainfall distribution with a rainfall deficit throughout the NER was observed during the 2018 monsoon (IMD, 2018 Southwest Monsoon End of Season Report). Thus, using 2017 provided the opportunity to examine the model behaviour in both low and high rainfall scenarios.

2.1 WRF daily accumulated rainfall

The Weather Research and Forecasting (WRF) model is a community based mesoscale model which is supported by the National Center for Atmospheric Research (NCAR) (Skamarock et al. 2008). The capabilities and faculties of WRF have been extensively used in many applications from atmospheric chemistry (Tie et al. 2010; Zhou et al. 2017), hydrology (Naabil et al. 2017; Srivastava et al. 2015; Zhao et al. 2009), lightning (Fierro et al. 2012; Giannaros et al. 2015, 2016; McCaul Jr. et al. 2009), and cyclones (Hsiao et al. 2012; Raju et al. 2011; Wang et al. 2011) to extreme rainfall events (Cardoso et al. 2013; Kumar et al. 2008). It has been developed with multiple dynamic and physics options. WRF houses two dynamical cores [advanced research WRF (ARW) and non-hydrostatic mesoscale model (NMM)]. WRF uses a nesting approach for simulating a high-resolution state of the atmosphere. It also facilitates 3-Dimensional and 4-dimensional VARiational data assimilation framework, namely, 3DVAR and 4DVAR, respectively. ARW-WRF version-3.5 is used here to simulate the daily accumulated rainfall augmented with its 3DVAR module to assimilate satellite products from global satellites through a two-way nesting approach (Fig. 1). The meteorological data used for the initialization of the model are obtained from the global forecasting system (GFS), a model data of half-degree grid resolution distributed by the national oceanic and atmospheric administration (NOAA). These initial conditions are then further augmented with data from the NCEP ADP global air and surface weather observation which are assimilated into it using 3DVAR. Land use land cover (LULC) data available with WRF model is replaced with ISRO NRSC (2012–2013) LULC data at a resolution of 30 s in this study. The model specifications for this study are described in Table 1.

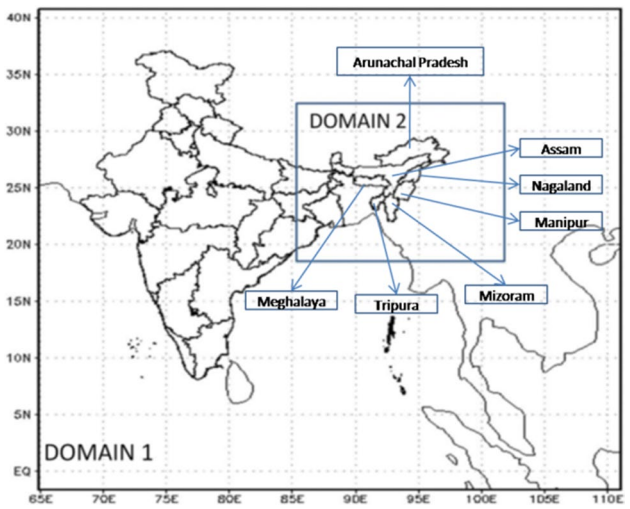

Fig. 1 Two-way nesting and the expanse of the domains (domain 1: 27 km; domain 2: 9 km)

2.2 GPM satellite rainfall estimate

The satellite rainfall used in this study is obtained from the global precipitation measurement (GPM) mission. The founding concept of the GPM mission is the utilisation of a constellation of precipitation observations contributed by national and international organisations to generate a temporally and spatially consistent rainfall estimate. Consequently, the GPM mission is divided into two aspects, one is the GPM core satellite (GPM-CO), and another is the GPM constellation of satellites. The GPM-CO, a NASA-JAXA partnership, was launched in February 2014. It operates in a non-sun-synchronous orbit with an inclination angle of 65°, giving it the ability of global coverage (Hou et al. 2014; Skofronick-Jackson et al. 2017). GPM-CO is an advanced successor to the tropical rainfall measurement mission (TRMM) that consists of a dual-frequency precipitation radar (DPR) (35.5 GHz and 13.6 GHz) and a 13-channel (10–183 GHz) conically scanning MW radiometer (GPM Microwave Imager: GMI). Both DPR and GMI can sense light rain and falling snow. DPR has a minimum detectable precipitation rate of 0.2 mm/h, indicating that it can detect the occurrence of 99% of the precipitation (Hamada and Takayabu 2016). TRMM's reach has been expanded by

Table 1 Details of the WRF simulation design

Spatial resolution	2 Domains: 27 kms (Indian sub-continent) and 9 kms (NER)
Temporal resolution	Daily
Data for initialization	Global forecast system (GFS)
Parameterization schemes	Micro-physics scheme = Thompson Long-wave radiation scheme = rrtm scheme Short-wave radiation scheme = Dudhia scheme Surface layer scheme = Monin–Obukhov (Janjic) scheme Land surface model = Unified Noah land-surface model PBL scheme = Mellor-Yamada-Janjic TKE scheme Cumulus parameterization scheme = Grell 3D ensemble scheme
Land-use-land-cover maps	National remote sensing centre (NRSC) updated land-use land cover maps

GPM, not only through its global coverage but also through inter-calibration of datasets from other MW imagers, and coordinated merged precipitation datasets (Integrated Multi-satellitE Retrievals for GPM: IMERG) (Huffman et al. 2015; He et al. 2017; Prakash et al. 2016). GMI brightness temperature is used by the IMERG algorithm to inter-calibrate data from the MW imagers in the GPM constellation. The temporal and spatial gaps in the MW precipitation estimates are filled by morphing the data in between the MW swaths and incorporating IR estimates using Kalman filter (KF). Thus producing a globally consistent product of $0.1° \times 0.1°$ and 30 min. resolution. The IR estimates are also calibrated by the MW data for an improved estimate of the same. The IMERG rainfall product consists of an MW-calibrated IR rainfall (GIR), a merged inter-calibrated MW rainfall (GMW), and an IMERG IR-MW merged rainfall (GMS; MS stands for multi-sensor). All three IMERG products with a spatial and temporal resolution of $0.1° \times 0.1°$ and 30 min, respectively, are used in the study. As the three IMERG products will have different characteristics, they will help in understanding the behaviour of the performance metrics under the influence of different error characteristics. The IMERG rainfall is available in three product versions, Early, late, and final. The IMERG Final run was observed to be the best among the three (Foelsche et al. 2017). The GMS rainfall estimate from the IMERG Final product has been bias-corrected using gauge rainfall from the global precipitation climatology centre (GPCC) (a detailed description on IMERG is available in its algorithm theoretical basis document (Huffman et al. 2015). IMERG version-5 is being used in this study, as its reliability has improved from its previous versions (Hou et al. 2014).

2.3 AWS rainfall

The AWS stations are installed and maintained at various sites in NER by the Indian Space Research Organization, Department of Space, Government of India. Five AWS stations (Dhemaji, Jorhat, Jowai, Tawang, and Udaypur) are chosen such that each station represents different regions, from valleys to mountains to cropland to temperate forests. The station locations with topography are presented in Fig. 2. The temporal consistency in the data during the analysis period is the criteria followed for selecting the stations. The station details are provided in Table 2.

3 Methodology

The workflow followed in this study and the performance metrics used are discussed in this section. Prior to the evaluation analysis, all the data products are pre-processed. The conversion of hourly or sub-hourly rainfall to daily

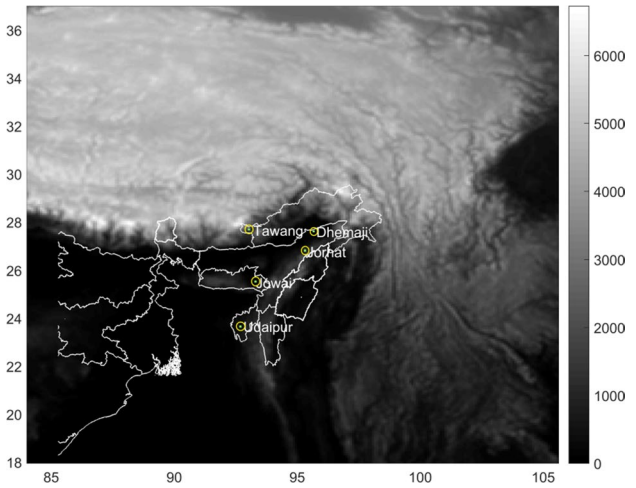

Fig. 2 AWS locations (Dhemaji, Jorhat, Jowai, Tawang, and Udaipur) are placed over a topography (meters) map

accumulated rainfall along with their spatial and temporal collocation is achieved during pre-processing. The workflow after pre-processing has been categorised into two stages. First is the evaluation of WRF with IMERG rainfall products and second is the analysis of WRF and IMERG rainfall products with AWS station rainfall. The efficacy of λ is examined in both the stages. The first stage begins with the analysis of the spatial distribution of mean seasonal rainfall and its standard deviation from all the rainfall products. The error analysis of WRF with IMERG is initially performed with the contemporary performance metrics: Bias, RMSD, and CC. Then, the characteristics of the new metric, in comparing WRF to IMERG, is analysed. A similar strategy is followed in the second stage.

3.1 Classic performance metrics

The performance metrics estimate the deviation and association between the target and reference dataset. The reference dataset is considered to be a representation of the truth. However, that is rarely the case, thus the notion of an RMSD is used here, rather than an RMSE (root mean square error), because none of the reference datasets are considered as the truth. The Bias (Eq. 1) would give a sense of the mean deviation between model and reference rainfall. The information from Bias is augmented with the standard deviation of residual (σ_R). The temporal variation of the difference between WRF and IMERG is given by σ_R, which would eventually lead to MSD (mean square deviation) or RMSD (Eq. 2c or 2f). The congruence between the temporal patterns of the two products is given by CC (Eq. 3):

Table 2 Details of AWS stations

Station name	Location	Meteorological subdivision	Altitude (m)	Temporal resolution	Terrain and climate characteristics
ISRO1104_15F450 (Dhemaji Degree College, Dhemaji)	Assam (27.47°, 94.55°)	Assam and Meghalaya	104	Hourly	Plain/warm temperate, winter dry, hot summer (Cwa; Köppen Geiger climate classification)
ISRO1068_15F42C (TRA, Jorhat)	Assam (26.73°, 94.23°)	Assam and Meghalaya	116	Hourly	Plain/warm temperate, winter dry, hot summer (Cwa; Köppen Geiger climate classification)
ISRO0042_15F02A (Jowai)	Meghalaya (25.425°, 92.17°)	Assam and Meghalaya	1300	Hourly	Hilly/Warm temperate, winter dry, warm summer (Cwb; Köppen Geiger climate classification)
ISRO0116_15F074 (Tawang)	Arunachal (27.59°, 91.87°)	Arunachal Pradesh	2669	Hourly	Mountainous/warm temperate, winter dry, warm summer (Cwb; Köppen Geiger climate classification)
ISRO0059_15F03B (Udaypur)	Tripura (23.51°, 91.48°)	Nagaland, Manipur, Mizoram, Tripura (NMMT)	120	Hourly	Plain/equatorial, winter dry (Aw; Köppen Geiger climate classification)

The climate characteristics are from the Köppen Geiger climate classification (Kottek et al. 2006)

$$\text{Bias} = n^{-1}\sum_{i=1}^{n}\left(x_i - y_i\right) = \bar{x} - \bar{y}, \tag{1}$$

$$\text{MSD} = n^{-1}\sum_{i=1}^{n}\left(x_i - y_i\right)^2, \tag{2a}$$

$$R_i = \left(x_i - y_i\right), \text{ where } i = 1 \text{ to } n, \tag{2b}$$

$$\text{MSD} = \text{Bias}^2 + \sigma_R^2, \tag{2c}$$

$$\sigma_R^2 = \sigma_x^2 + \sigma_y^2 - 2\text{Cov}\left[x, y\right], \tag{2d}$$

$$\text{MSD} = \text{Bias}^2 + \sigma_x^2 + \sigma_y^2 - 2\text{Cov}\left[x, y\right], \tag{2e}$$

$$\text{RMSD} = \sqrt{\text{MSD}}, \tag{2f}$$

$$\text{CC} = \frac{\text{Cov}\left[x, y\right]}{\sqrt{\sigma_x^2 \times \sigma_y^2}}, \tag{3}$$

where "x" is WRF rainfall and "y" is IMERG rainfall products, and while comparing WRF and IMERG with AWS rainfall, "y" represents AWS and "x" represents WRF or IMERG products. The standard deviation of x is given by σ_x, and that of y is represented by σ_y. The residual, the deviation between x and y, is referred as R. The correlation is computed only for statistically significant points ($p < 0.05$), obtained using Student T test. As is shown by Eq. (2e, f), the behaviour of *RMSD* is governed by Bias, σ_x, σ_y, and $\text{Cov}\left[x, y\right]$. Thus, prior to an analysis of RMSD, the spatial map of σ_x and σ_y are analysed to better understand its characteristics.

3.2 Symmetric index of agreement

The desired properties of an index of agreement are dimensionless, bounded, symmetric, easy to compute, and interpretable to a standard measure of performance, like correlation (Duveiller et al. 2016). Many indices of agreement have joined the fray that mostly agreed with these properties, like Nash–Sutcliffe Efficiency (Nash and Sutcliffe 1970), index of agreement (Willmott 1981), and refined indices of agreement (Mielke 1984, 1991; Willmott et al. 2012). However, these indices do not consider symmetry as a property in their formulation, i.e., they consider a reference data as a representation of the truth. Thus, if X is considered as the reference data, an index comparing X and Y will not be the same as an index for Y and X. But, in case of an inter-comparison study, where it not required knowing before-hand which of the data sets is better or closer to the truth, it is difficult to identify a reference data. Thus, in this case it is required for the index comparing X and Y to be same as an index comparing Y and X (Duveiller et al. 2016). This is the property of symmetry.

Considering the desirable properties of an index of agreement, Duveiller et al. (2016) set out to develop an index that encompasses all of these properties based on the basic formulation given in Eq. (4) (Willmott et al. 2012):

$$\rho = 1 - \frac{\delta}{\mu}, \tag{4}$$

where δ represents the deviation between the target dataset and the reference observation (e.g., MSD) such that it is always greater than or equal to zero ($\delta \geq 0$), and μ is a set of reference deviations. This formulation should be devised in such a manner that, $\mu > \delta$, so that the value of ρ never exceeds beyond 1. Thus, the upper bound of ρ is 1, which indicates a perfect similarity between the two datasets. The lower bound of ρ is assigned values of either 0, −1, or $-\infty$ depending on the different formulations of Eq. (4). A poor agreement is indicated by the lower bound of ρ. Mielke (1984, 1991) proposed the first symmetric index of agreement which is represented as

$$\delta = n^{-1} \sum_{i=1}^{n} \left(x_i - y_i\right)^{\gamma}, \tag{5a}$$

$$\mu = n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} \left(x_i - y_j\right)^{\gamma}, \tag{5b}$$

$$\rho = 1 - \frac{n^{-1} \sum_{i=1}^{n} \left(x_i - y_i\right)^{\gamma}}{n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} \left(x_i - y_j\right)^{\gamma}}, \tag{5c}$$

where μ is the sum of the differences between each time point of "X" with every time other point of "Y" (Eq. 5b) and δ shows the sum of the difference between X and Y at each time (Eq. 5a). One of the disadvantages of this formulation of ρ by Mielke (Eq. 5c) is that the estimation of μ is computationally expensive for large "n". However, for $\gamma = 2$, δ becomes same as MSD from Eq. (2e) and μ can be simplified in terms of σ_x^2, σ_y^2, and Bias (Eq. 7) (this simplification of μ is available in the supplementary information of Duveiller et al. 2016). By replacing the new formulation of μ and δ in Eq. (5c), ρ becomes as in Eq. (8).

$$\delta = \text{Bias}^2 + \sigma_x^2 + \sigma_y^2 - 2\text{Cov}\left[x, y\right], \tag{6}$$

$$\mu = \text{Bias}^2 + \sigma_x^2 + \sigma_y^2, \tag{7}$$

$$\rho = 1 - \frac{\text{Bias}^2 + \sigma_x^2 + \sigma_y^2 - 2\text{cov}\left[x, y\right]}{\text{Bias}^2 + \sigma_x^2 + \sigma_y^2}. \tag{8}$$

However, this formulation of ρ (Eq. 8) is not upholding the criteria of $\mu > \delta$. This issue is resolved by adding a co-variance term (κ), that it is always positive, to μ from Eq. (7).

$$\kappa = 2\left|\text{Cov}\left[x, y\right]\right|. \tag{9a}$$

$$\mu = \text{Bias}^2 + \sigma_x^2 + \sigma_y^2 + \kappa. \tag{9b}$$

It was observed by Duveiller et al. (2016) that by adding κ separately to μ (Eq. (7)), it ensures that $\mu = \delta$ when $\text{CC} < 0$, and thus makes $\rho = 0$. But when $\text{CC} > 0$, ρ seems to be inflated as δ is always less than μ. It is because of the negative sign in front of the co-variance term in δ (Eq. 6). To solve this issue a condition is incorporated in the definition of κ shown in Eq. (10). The final form of ρ is shown in Eq. (11), where ρ is replaced by the notation λ to avoid any confusion. Here on out λ will refer to the symmetric index of agreement and will be defined by Eq. (11).

$$\kappa = \begin{cases} 0; \text{CC} \geq 0 \\ 2\left|\text{Cov}\left[x, y\right]\right|; \text{CC} < 0. \end{cases} \tag{10}$$

$$\lambda = 1 - \frac{\text{Bias}^2 + \sigma_x^2 + \sigma_y^2 - 2\text{cov}\left[x, y\right]}{\text{Bias}^2 + \sigma_x^2 + \sigma_y^2 + \kappa}. \tag{11}$$

From Eqs. (4), (8), and (10), the final form of Eq. (11) is obtained. For $\text{CC} \geq 0$: $\kappa = 0$, λ from Eq. (11) can be reformulated in terms of correlation (CC) as:

$$\lambda = \alpha \cdot \text{CC}. \tag{12a}$$

$$\alpha = \frac{2}{\frac{\sigma_x}{\sigma_y} + \frac{\sigma_y}{\sigma_x} + \frac{\text{Bias}^2}{(\sigma_x \sigma_y)}}. \tag{12b}$$

Equation (12a, b) is only valid for $\text{CC} \geq 0$ and considers $\lambda = 0$ when $\text{CC} < 0$. From Eq. (12), it is evident that when Bias ihigh, it weighs CC down, lowering the value of λ. Although this relation between CC and Bias is not linear but quadratic, i.e., if the Bias is high, it has a significant influence on CC. However, if it is marginal, its influence on CC is insignificant. This behaviour is because Bias in Eq. (12) is a squared quantity. Thus, λ weighs CC based on Bias, thereby incorporating the property of "agreement" in temporal pattern and magnitude of deviation into one index. For $\text{CC} \geq 0$ and when there are no multiplicative and additive bias, $\lambda = \text{CC}$ (Duveiller et al. 2016), i.e., α becomes equal to 1 in Eq. (12a). Thus the advantage of this property is that the index value can be immediately compared to CC. Any deviation from *CC* indicates an increase in bias proportional to α. This formulation of λ is bounded between 0 (no agreement) and 1 (total agreement). The examination of Eq. (10) to Eq. (12) shows that λ can capture all the desired

properties of an index of agreement: dimensionless, bounded between 0 and 1, symmetric, easy to compute, and interpretable to a widely used performance measure, CC.

4 Results and discussion

The results from the comparative analysis between WRF simulated daily accumulated rainfall and GPM rainfall products (GIR, GMW, and GMS) are discussed here along with their comparison with respect to the AWS rainfall.

4.1 Evaluation of WRF with IMERG rainfall

The spatial map of JJAS seasonal mean rainfall and its standard deviation of GMS, GIR, GMW, and WRF are shown in Figs. 3 and 4, respectively. It is evident that the mean rainfall is concentrated in the South of NER and the East (Fig. 3). Both the model and the satellite are able to show the climatologically wet region of Meghalaya. Figure 3 also show that the northern most state of Arunachal and central state of Assam received very low seasonal rainfall compared to Manipur, Mizoram, Tripura and Nagaland. This coincides with the Monsoon Report of IMD regarding the regional heterogeneity of rainfall in 2017 (IMD, 2017 southwest monsoon end of season report). However, a notable difference in the magnitude of seasonal mean rainfall of GMW is observed with a maximum of ≈ 10 mm/day compared to a maximum of ≈ 40 mm/day in GMS, GIR, and WRF. Furthermore, it is also observed in Figs. 3 and 4 that the regions with a high seasonal mean rainfall correspond to regions with high seasonal variability. The similarity in the seasonal mean and its standard deviation between WRF and IMERG products infer a similarity in their temporal behaviours. However, the magnitude of rainfall varies from product to product. These deviations are captured by the performance metrics.

The spatial map of Bias in WRF when it is compared with the IMERG products (GMS, GIR, and GMW) and their corresponding σ_R is shown in Fig. 5. The Bias between WRF and IMERG is lowest for GMS and highest for GMW (Fig. 5a). Moreover, the Bias is similar when WRF is compared with GMS and GIR. This behaviour of Bias is because of the difference in their seasonal mean rainfall (Fig. 3). Unlike Bias, σ_R is low in case of a comparison between WRF and GMW and high in case of a comparison with GIR and GMS (Fig. 5b). A low σ_R indicates that WRF is consistently over-estimating or under-estimating the rainfall and a high σ_R indicates that the magnitude of deviation of WRF varies with time. The spatial map of σ_R in Fig. 5b is very similar to the corresponding spatial map of standard deviation of seasonal mean of WRF and IMERG products. This behaviour

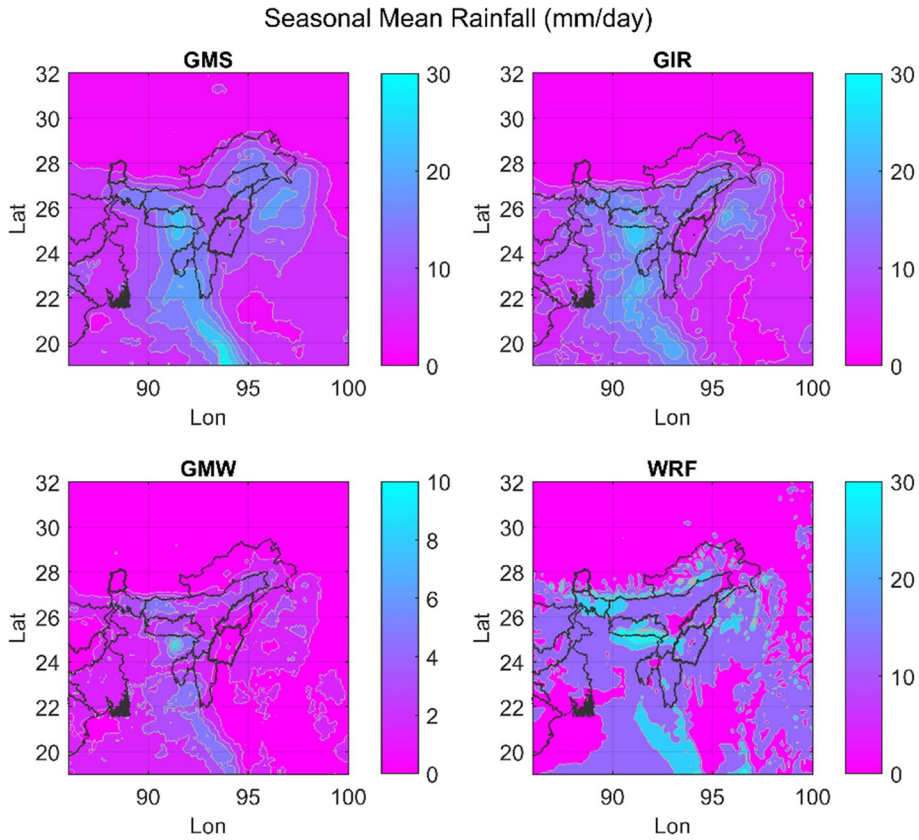

Fig. 3 JJAS seasonal mean of daily accumulated rainfall (mm/day) of GMS, GIR, GMW, and WRF

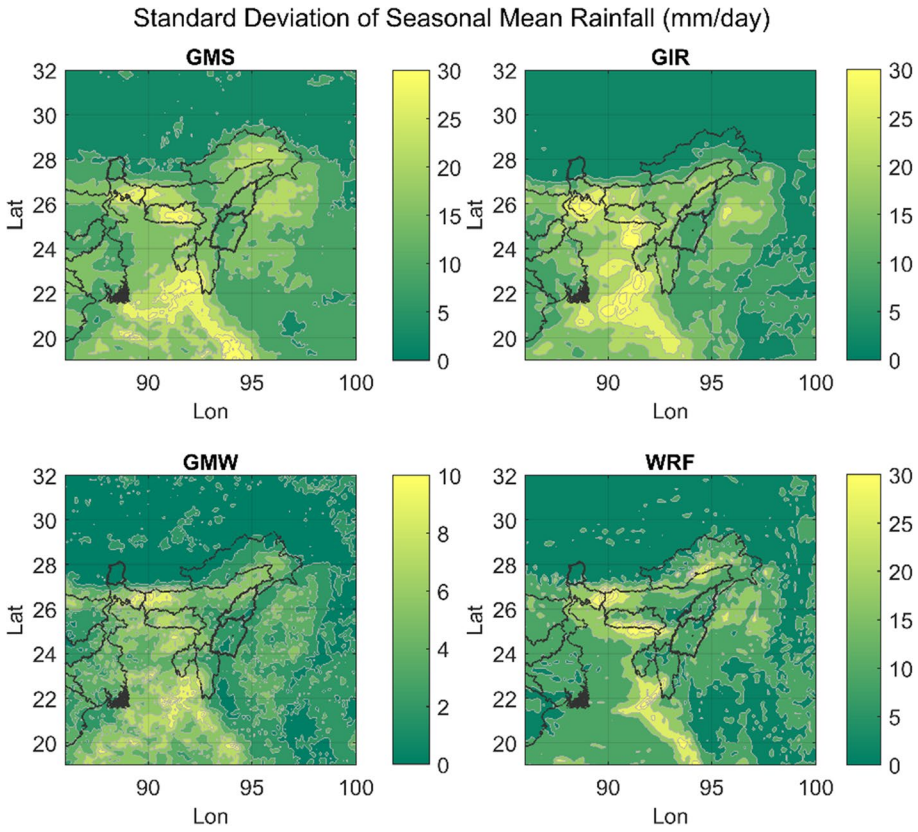

Fig. 4 Standard deviation of JJAS seasonal mean rainfall (mm/day) from GMS, GIR, GMW, and WRF

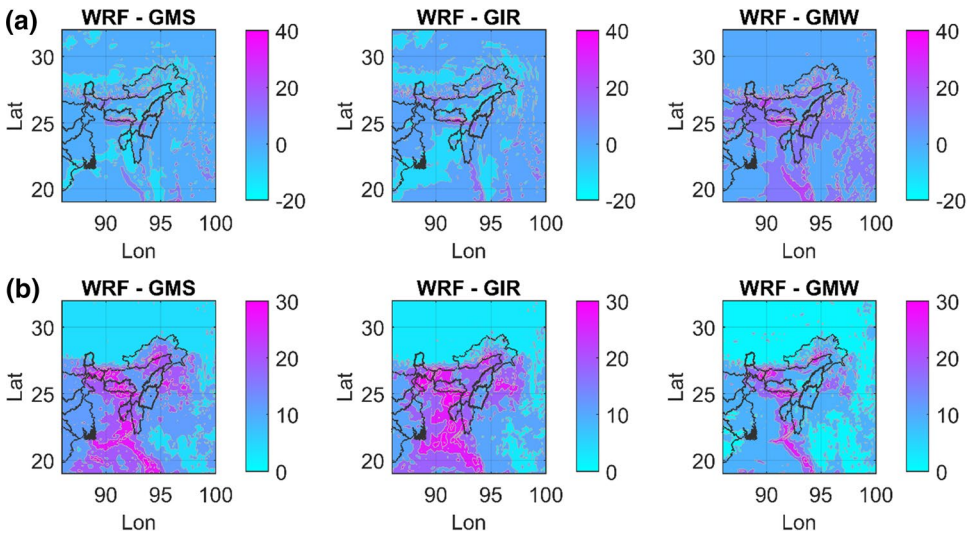

Fig. 5 **a** Bias in JJAS daily accumulated rainfall (mm/day) between WRF and GMS/GIR/GMW, **b** Standard deviation of residual (σ_R) of the JJAS daily accumulated rainfall (mm/day) between WRF and GMS/GIR/GMW

can be explained by the fact that σ_R can be decomposed into the standard deviations of seasonal mean and covariance (Eq. 2d). The extent of σ_R was essential to visualize, because it would help in better understanding the behaviour of RMSD.

The spatial map of RMSD and CC in NER is shown in Fig. 6. It is evident from Fig. 6a that WRF is very close to the IMERG rainfall products. The RMSD of WRF with IMERG products is observed to be ≈ 30 mm/day in most of the pixels (Fig. 6a). The valley region of Assam shows the lowest RMSD, and it increases with increase in topographic complexity. A patch of strong correlation (> 0.8) in the northeast of NER which coincides with low RMSD is evident in Fig. 6b. Apart from this, the distribution of CC over NER shows a relatively weak agreement in temporal patterns. Two fascinating features are apparent from Fig. 6.

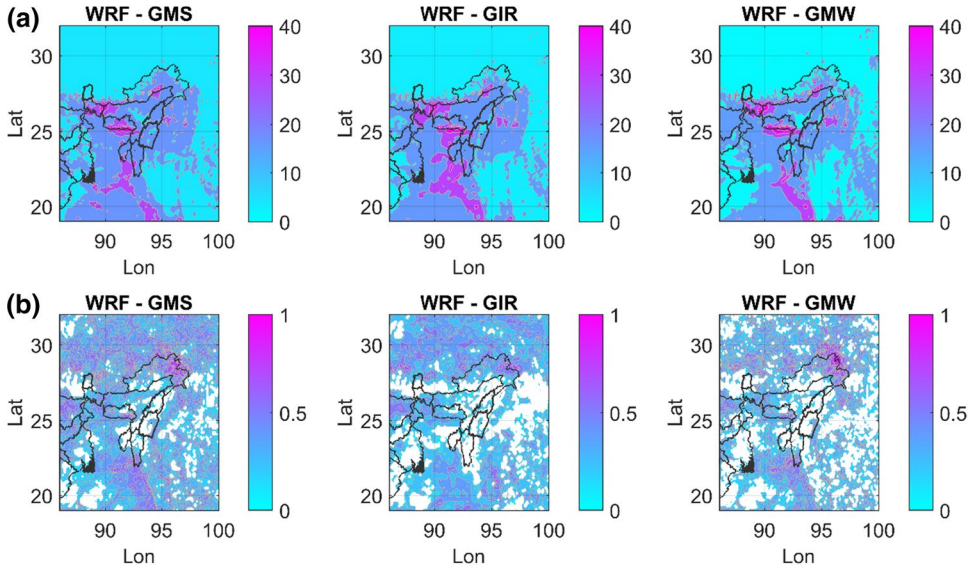

Fig. 6 **a** Spatial map of RMSD (mm/day) of WRF vs. GMS/GIR/GMW, **b** Spatial map of CC of WRF vs. GMS/GIR/GMW. Blank pixels in **b** represents areas, where CC is statistically not significant

First, some regions, like southern Meghalaya, have both high RMSD and high CC. Second is the similarity in the behaviour of RMSD and CC of WRF to IMERG products. This similarity is misleading, considering the strong Bias between WRF and GMW. The similarity of RMSDs can be explained by the inversely related behaviour of Bias and σ_R (Fig. 5), i.e., high Bias corresponds to low σ_R (WRF–GMW) and vice-versa (WRF–GMS and WRF–GIR). This explanation is evident from Eq. (2c) too. These behaviours of RMSD and CC are misleading and can lead to confusion in intercomparison studies.

The formulation of λ is expected to rectify the misleading aspects of the standard performance metrics. Figure 7 shows the behaviour of λ for WRF with respect to the IMERG rainfall products over NER. It shows that WRF has a relatively good agreement with GMS and GIR compared to GMW, which was not the case, according to RMSD and CC. It is so because λ is a function of Bias and not RMSD (Eq. 12). Another critical point to be noted from Figs. 6 and 7 is that regions with low RMSD and a strong correlation show high values of λ. However, the ambiguous regions, with high RMSD and strong correlation, are represented by low values of λ. These features of RMSD, CC, and λ are further discussed in Fig. 8 using histograms of the same. Figure 8a, b shows that when WRF is compared with the three reference datasets (GMS, GIR, and GMW), they provide similar results of percentage grid points covered by bins of RMSD and CC. However, Fig. 8c clearly shows the distinct performances of WRF with different reference dataset. Considering the debate over the efficacy of mean absolute deviation (MAD) or RMSD as an appropriate measure of deviation (Chai and Draxler 2014; Willmott and Matsuura 2005; Willmott et al. 2009), λ is also compared with MAD. It is observed that MAD behaves very similar to RMSD, showing similar agreement between WRF and the IMERG products, unlike λ (Fig. 9). To further analyse the characteristics of λ, WRF and IMERG products are compared against AWS rainfall using both the classic metrics and λ.

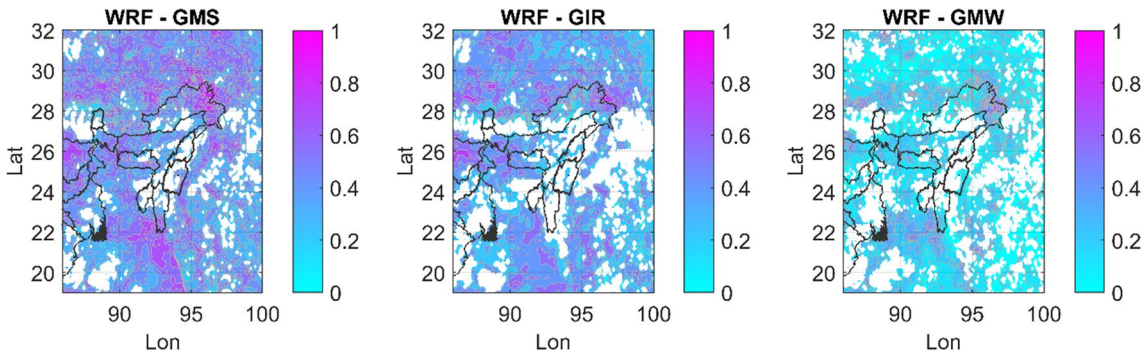

Fig. 7 Spatial map of λ of WRF vs. GMS, GIR, and GMW. Blank pixels represent areas, where CC is statistically not significant ($p < 0.05$)

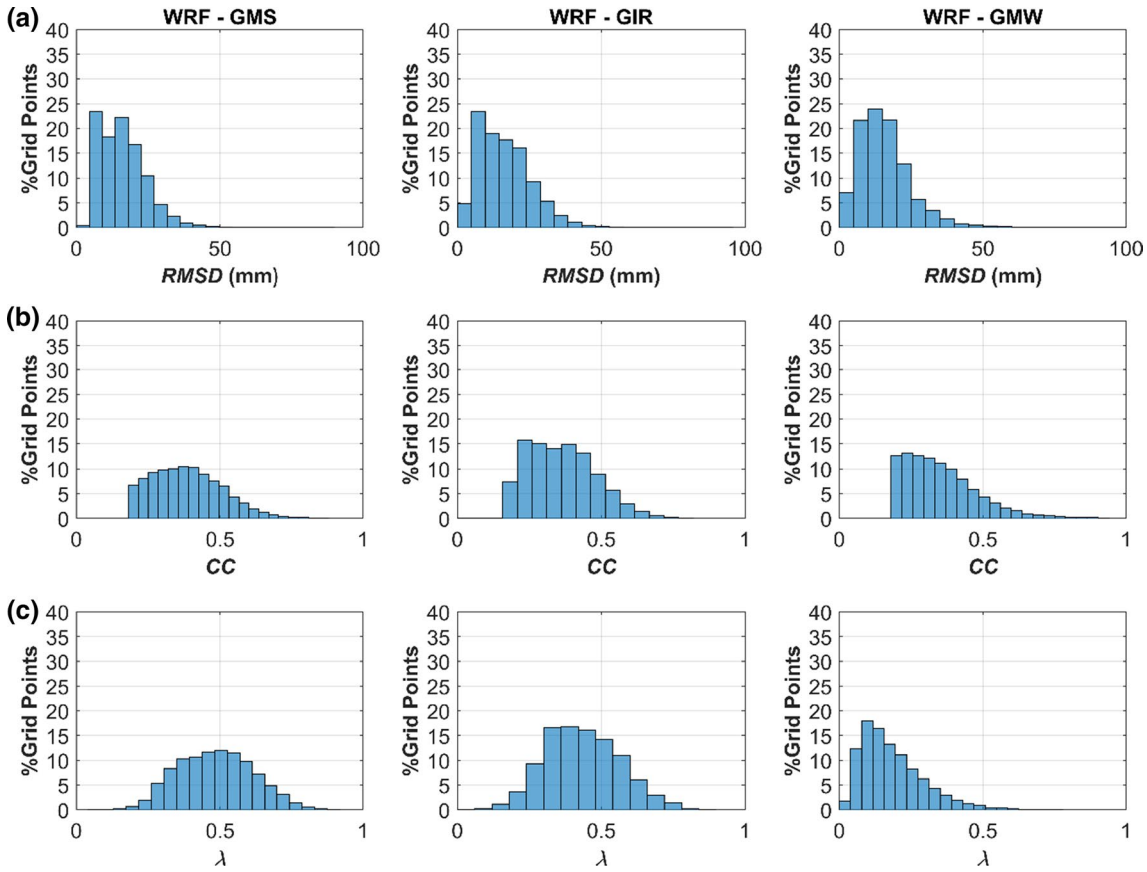

Fig. 8 **a** Histogram of percentage grid points under bins of RMSD (mm) of WRF vs. GMS/GIR/GMW, **b** Histogram of percentage grid points under bins of CC of WRF vs. GMS/GIR/GMW, **c** Histogram of percentage grid points under bins of λ of WRF vs. GMS/GIR/GMW

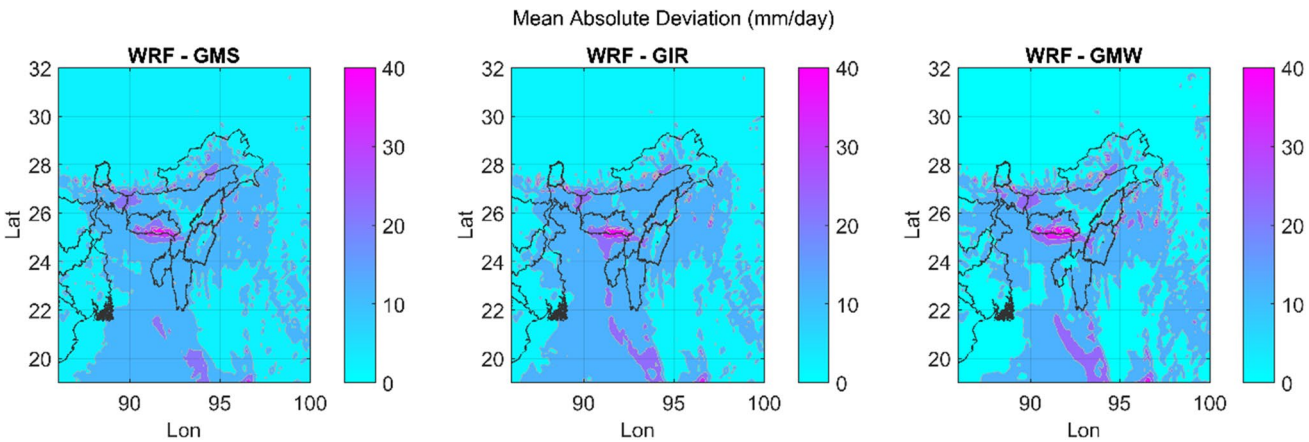

Fig. 9 Spatial map of MAD (mm/day) of WRF vs. GMS/GIR/GMW

4.2 Evaluation of WRF and IMERG with AWS rainfall

Figure 10 shows the inter-comparison of WRF and IMERG products with respect AWS rainfall at five locations across NER (Dhemaji, Jorhat, Jowai, Tawang, and Udaipur). GMS shows better agreement with AWS compared to other IMERG products. The strongest agreement between GMS and AWS is observed at Jowai (RMSD: 76 mm, CC: 0.75) and Udaipur (RMSD: 12.9 mm, CC: 0.84) and the weakest at Jorhat (RMSD: 20.5 mm, CC: 0.32). WRF, on the other

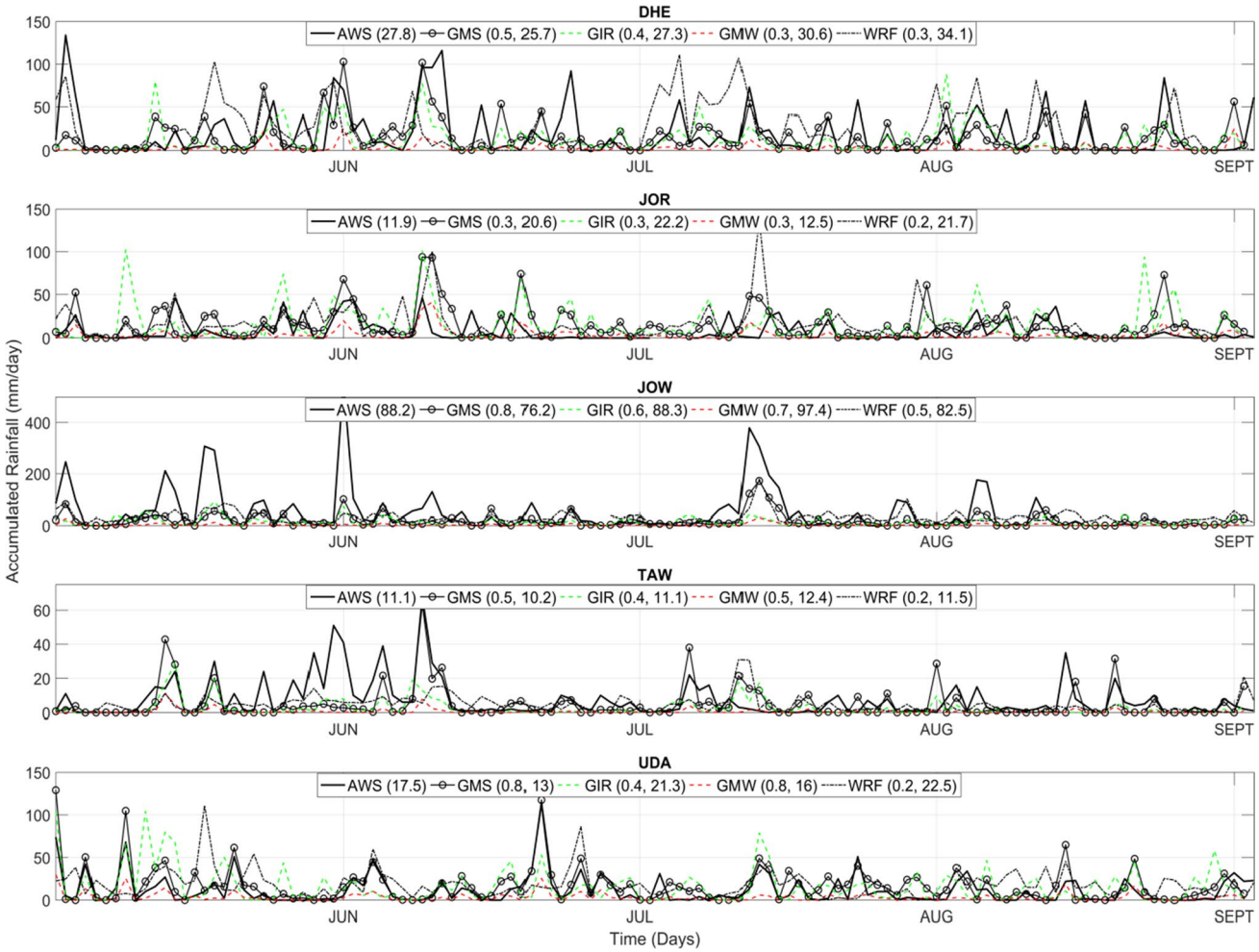

Fig. 10 Inter-comparison of WRF, GMS, GIR, and GMW daily accumulated rainfall with respect to AWS rainfall at 5 stations: Dhemaji (DHE), Jorhat (JOR), Jowai (JOW), Tawang (TAW), and Udaipur (UDA). The values in brackets for GPM products and WRF represent the respective CC and RMSD with respect to AWS, and the bracket adjacent to AWS represents its standard deviation

hand, showed relatively weak agreement with AWS, compared to GPM. The strongest agreement is at Jowai (RMSD: 82 mm, CC: 0.47) and the weakest at Jorhat (RMSD: 21.7 mm, CC: 0.17):

$$\text{NBias} = \frac{\sum_{i=1}^{n}\left(x_i - y_i\right)}{\sum_{i=1}^{n} y_i}. \tag{13}$$

where "y" represents AWS and "x" represents WRF or IMERG products.

Since RMSD and CC can lead to ambiguous interpretations (Figs. 6, 7, 8), the performance of WRF and IMERG against AWS is also investigated using λ. This analysis also uses Bias and CC to investigate the characteristics of their influence on λ. Figure 11 shows the behaviour of λ to Normalized-Bias (Nbias) and CC for different datasets and at different stations. NBias is being used here to bring the scale of Bias at par with CC and λ (Eq. 13; Kryza et al. 2010). Figure 11 shows that the impact of Bias on CC, as reflected in λ, is not linear, i.e., it does not give equal weights to all the deviations. If the Bias is large enough, then only its influence on CC would be visible in λ. For example, although Nbias of GMS at Jorhat is high, because CC is already low, there has not been much impact on it as reflected in λ. However, consider the case at Jowai. Since the value of CC is significantly large enough, even a marginally high Nbias caused the CC values to go down a lot, as is reflected in λ. Also based on the analysis of λ, it can be said with appreciable certainty that GMS is a relatively good dataset to be used in plains, like Udaipur, and complex topographic regions, like Tawang. The comparatively poor performance of GMW can be associated with the performance of its member products (Figs. 7, 11). Similar behaviour has been observed by Tan et al. (2016), and they further emphasized that the behaviour of GMW is not because of its morphing algorithm. So as GMS and

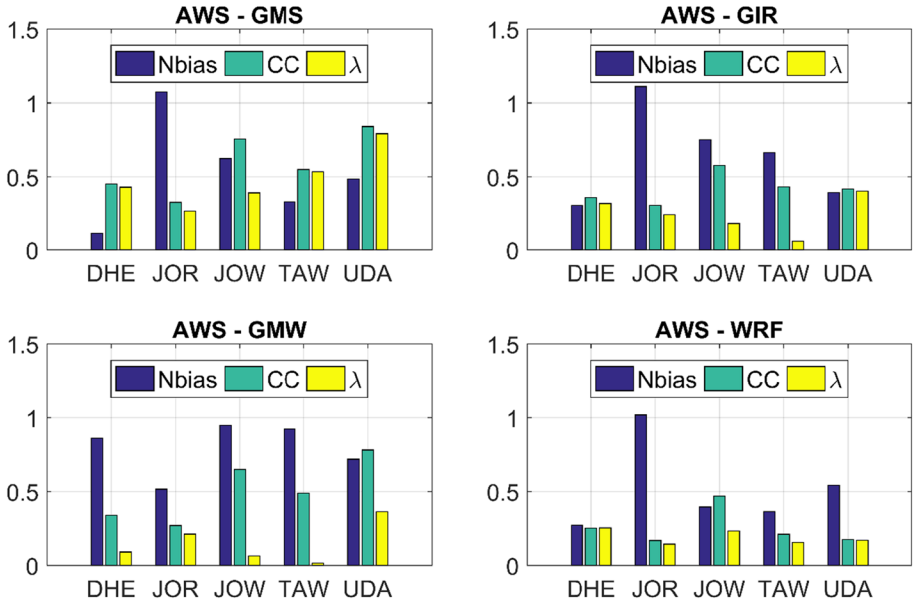

Fig.11 Inter-comparison of λ with respect to Nbias and CC estimated for WRF, GMS, GIR, and GMW with respect to AWS rainfall at five stations [Dhemaji (DHE), Jorhat (JOR), Jowai (JOW), Tawang (TAW), and Udaipur (UDA)]

GIR incorporate GMW, it also inherits its errors (Prakash et al. 2016).

5 Summary and conclusion

This study had two aspects to cater to, first is to investigate the applicability of a new performance metric in a comparative analysis and second is the behaviour of WRF simulated daily accumulated rainfall against IMERG rainfall products (GMS, GIR, and GMW) and AWS daily accumulated rainfall. IMERG datasets and AWS rainfall acted as reference datasets for evaluating WRF. Along with the standard performance metrics used in this study [Bias, root mean square deviation (RMSD), and correlation coefficient (CC)], a new performance metric namely, the symmetric index of agreement (λ) was also applied. This index reflects the characteristics of both similarity in temporal patterns (CC) and magnitude of deviation (Bias). The rationale behind the introduction of λ is to examine its ability as a holistic measure of skill and rectify the limitations of RMSD along the way.

The investigation suggested that WRF rainfall had strong similarities with GMS and GIR than GMW. A comparison between WRF and GMW using Bias and standard deviation of residual (σ_R) showed a high Bias and a low σ_R. However, a contradictory result was obtained for the comparison with GMS and GIR, low Bias and high σ_R. This behaviour of error was not reflected in the analysis using RMSD and CC, which showed similar values of RMSD and CC for all three IMERG datasets. It was attributed to the inverse relationship between Bias and σ_R, as RMSD is a function of both. Another significant finding of the analysis of standard performance metrics was that some areas with high RMSD corresponded with areas with high CC. These aspects of the behaviour of RMSD and CC can lead to an ambiguous conclusion. However, the application of λ for evaluating the performance of WRF showed that it could handle these aspects and generate an evaluation devoid of any ambiguity. The analysis using λ showed that WRF with GMW had the lowest values of λ compared to GIR and GMS. Thus, it showed that the influence of Bias is appropriately incorporated in λ. The ambiguous regions with high Bias and high CC had also been aptly represented with low values of λ.

The performance evaluation of IMERG products with AWS data showed that GMS was the most skilled, followed by GIR and GMW. This behaviour can be associated with the inclusion of poor MW products, when under the influence of topography. The enhanced performance of GMS can be attributed to its calibration by GPCC gauged rainfall. The analysis between WRF and AWS showed its relatively poor skill in representing the station rainfall. The inter-comparison analysis using IMERG products and WRF against AWS was also used to understand the relationship between Bias, CC, and λ. It showed that the influence of the magnitude of deviation on CC was not linear. Thus, when Bias is high, and CC is low, its impact on λ would also be low. Also, when Bias is high, and CC is also high, the impact on λ would also be high.

This study showed the inherent limitations of RMSD and CC and also showed the efficacy of λ in evaluating numerical weather prediction models. One of the applications of a single reliable measure of model skill is in model intercomparisons and sensitivity studies using multi-criteria decision analysis (Sikder and Hossain 2016). The evaluation of WRF showed that when compared to GMS, it showed a

good performance, but when compared to AWS, it was not reliable. It can be associated with the representation error between a point observation (AWS) and WRF. Furthermore, it also showed that a more extensive study of the effect of topography and the effect of planetary boundary layer (PBL) on WRF rainfall simulations is required.

Since studies have suggested that the significance of spatial information has increased in the validation of high-resolution rainfall forecasts (Gilleland et al. 2009), many new spatial verification measures have been developed. For example, Fractions Skill Score: FSS (Roberts and Lean 2008) and Structure, Amplitude, and Location: SAL (Wernli et al. 2008) to assess the accuracy of structure and location of simulated rainfall events. However, it is pertinent to mention here that the classic performance metrics that this study deals with are used for point-to-point validation with no regard for spatial information. Since RMSE, CC, and Bias do not take into account spatial information like, structure and location of an event, it can be hypothesized that λ also will not accommodate these information. However, the spatial information in validation studies refers not only to location and structure of a rainfall event but also to the effect of scale on the forecast error. Thus, one of the ways λ can accommodate the spatial information is by identifying the effect of spatial scales on forecast error. This can be achieved by decomposing the variable field into various scale components using 2D-wavelet analysis or a 2D-singular spectrum analysis (2D-SSA) and applying λ onto the decomposed scale components. This type of an analysis has already been implemented for RMSD and CC to investigate the effect the scale on forecast error (Briggs and Levine 1997; Casati et al. 2004). This analysis was beyond the scope of this study, but can be taken up in the future.

Acknowledgements The authors would like to acknowledge the North Eastern Space Applications Centre for providing the infrastructure required to undertake this work. The authors would also like to thank Ms. Ritu Anil Kumar for her valuable suggestions during the manuscript formulation process. The authors would also like to extend their gratitude to the anonymous reviewers for their constructive suggestions in improving the presentation of the manuscript.

References

Abbe C (1901) The physical basis of long-range weather forecasts. Mon Weather Rev. https://doi.org/10.1175/1520049 3(1901)29[551c:TPBOLW]2.0.CO;2

Adnan M, Khan F, Rehman N, Ali S, Hassan SS, Dogar MM et al (2020) Variability and predictability of summer monsoon rainfall over Pakistan. Asia-Pacific J Atmos Sci 1–9. https://doi.org/10.1007/s13143-020-00178-2

Bauer P, Thorpe A, Brunet G (2015) The quiet revolution of numerical weather prediction. Nature. https://doi.org/10.1038/nature14956

Bharti V, Singh C (2015) Evaluation of error in TRMM 3B42V7 precipitation estimates over the Himalayan region. J Geophys Res Atmos 120(24):12458–12473. https://doi.org/10.1002/2015JD023779

Bjerknes V (1910) Dynamic meteorology and hydrography (No. 88). Carnegie.

Cardoso RM, Soares PMM, Miranda PMA, Belo-Pereira M (2013) WRF high resolution simulation of Iberian mean and extreme precipitation climate. Int J Climatol. https://doi.org/10.1002/joc.3616

Carpenter KM (1979) An experimental forecast using a non-hydrostatic mesoscale model. Quart J R Meteorolog Soc. https://doi.org/10.1002/qj.49710544510

Chai T, Draxler RR (2014) Root mean square error (RMSE) or mean absolute error (MAE)?—arguments against avoiding RMSE in the literature. Geosci Model Dev. https://doi.org/10.5194/gmd-7-1247-2014

Chawla I, Osuri KK, Mujumdar PP, Niyogi D (2018) Assessment of the weather research and forecasting (WRF) model for simulation of extreme rainfall events in the upper Ganga Basin. Hydrol Earth Syst Sci. https://doi.org/10.5194/hess-22-1095-2018

Dudhia J (2014) A history of mesoscale model development. Asia-Pacific J Atmos Sci. https://doi.org/10.1007/s13143-014-0031-8

Duveiller G, Fasbender D, Meroni M (2016) Revisiting the concept of a symmetric index of agreement for continuous datasets. Sci Rep. https://doi.org/10.1038/srep19401

Ebert EE, Janowiak JE, Kidd C (2007) Comparison of near-real-time precipitation estimates from satellite observations and numerical models. Bull Am Meteor Soc. https://doi.org/10.1175/BAMS-88-1-47

Fierro AO, Mansell ER, Ziegler CL, MacGorman DR (2012) Application of a lightning data assimilation technique in the WRF-ARW model at cloud-resolving scales for the tornado outbreak of 24 May 2011. Mon Weather Rev 140(8):2609–2627. https://doi.org/10.1175/MWR-D-11-00299.1

Foelsche U, Kirchengast G, Fuchsberger J, Tan J, Petersen WA (2017) Evaluation of GPM IMERG early, late, and final rainfall estimates using WegenerNet gauge data in southeastern Austria. Hydrol Earth Syst Sci 21(12):6559–6572. https://doi.org/10.5194/hess-21-6559-2017

Ghile YB, Schulze RE (2010) Evaluation of three numerical weather prediction models for short and medium range agrohydrological applications. Water Resour Manage 24(5):1005–1028. https://doi.org/10.1007/s11269-009-9483-5

Giannaros TM, Melas D, Daglis IA, Keramitsoglou I, Kourtidis K (2013) Numerical study of the urban heat island over Athens (Greece) with the WRF model. Atmos Environ 73:103–111. https://doi.org/10.1016/j.atmosenv.2013.02.055

Giannaros TM, Kotroni V, Lagouvardos K (2015) Predicting lightning activity in Greece with the weather research and forecasting (WRF) model. Atmos Res 156:1–13. https://doi.org/10.1016/j.atmosres.2014.12.009

Giannaros TM, Kotroni V, Lagouvardos K (2016) WRF-LTNGDA: a lightning data assimilation technique implemented in the WRF model for improving precipitation forecasts. Environ Model Softw 76:54–68. https://doi.org/10.1016/j.envsoft.2015.11.017

Gilleland E, Ahijevych D, Brown BG, Casati B, Ebert EE (2009) Intercomparison of spatial forecast verification methods. Weather Forecast 24(5):1416–1430. https://doi.org/10.1175/2009WAF2222269.1

Hamada A, Takayabu YN (2016) Improvements in detection of light precipitation with the global precipitation measurement dual-frequency precipitation radar (GPM DPR). J Atmos Ocean Technol 33(4):653–667. https://doi.org/10.1175/JTECH-D-15-0097.1

He Z, Yang L, Tian F, Ni G, Hou A, Lu H (2017) Intercomparisons of rainfall estimates from TRMM and GPM multisatellite products over the upper Mekong River basin. J Hydrometeorol 18(2):413–430. https://doi.org/10.1175/JHM-D-16-0198.1

Hou AY, Kakar RK, Neeck S, Azarbarzin AA, Kummerow CD, Kojima M, Oki R, Nakamura K, Iguchi T (2014) The global precipitation measurement mission. Bull Am Meteor Soc 95(5):701–722. https://doi.org/10.1175/BAMS-D-13-00164.1

Hsiao LF, Chen DS, Kuo YH, Guo YR, Yeh TC, Hong JS, Fong CT, Lee CS (2012) Application of WRF 3DVAR to operational typhoon prediction in Taiwan: impact of outer loop and partial cycling approaches. Weather Forecast 27(5):1249–1263. https://doi.org/10.1175/WAF-D-11-00131.1

Huffman GJ, Bolvin DT, Braithwaite D, Hsu K, Joyce R, Xie P, Yoo SH (2015) NASA global precipitation measurement (GPM) integrated multi-satellite retrievals for GPM (IMERG). Algorithm Theoretical Basis Document Version4.5. https://pmm.nasa.gov/sites/default/files/document_files/IMERG_ATBD_V4.5.pdf

Jiménez PA, Dudhia J (2013) On the ability of the WRF model to reproduce the surface wind direction over complex terrain. J Appl Meteorol Climatol 52(7):1610–1617. https://doi.org/10.1175/JAMC-D-12-0266.1

Kalnay E, Kanamitsu M, Baker WE (1990) Global numerical weather prediction at the National Meteorological Center. Bull Am Meteor Soc 71(10):1410–1428. https://doi.org/10.1175/1520-0477(1990)071%3c1410:GNWPAT%3e2.0.CO;2

Kotroni V, Lagouvardos K, Retalis A (2011) The heat wave of June 2007 in Athens, Greece-Part 2: modeling study and sensitivity experiments. Atmos Res 100(1):1–11. https://doi.org/10.1016/j.atmosres.2010.12.007

Kottek M, Grieser J, Beck C, Rudolf B, Rubel F (2006) World map of the Köppen-Geiger climate classification updated. Meteorolog Z 15(3):259–263. https://doi.org/10.1127/0941-2948/2006/0130

Kryza M, Szymanowski M, Migała K, Pietras M (2010) Spatial information on total solar radiation: application and evaluation of the r.sun model for the Wedel Jarlsberg Land, Svalbard. Pol Polar Res 31(1):17–32. https://doi.org/10.4202/ppres.2010.02

Kryza M, Werner M, Wałszek K, Dore AJ (2013) Application and evaluation of the WRF model for high-resolution forecasting of rainfall—a case study of SW Poland. Meteorolog Z 22(5):595–601. https://doi.org/10.1127/0941-2948/2013/0444

Kumar A, Dudhia J, Rotunno R, Niyogi D, Mohanty UC (2008) Analysis of the 26 July 2005 heavy rain event over Mumbai, India using the weather research and forecasting (WRF) model. Quart J R Meteorolog Soc 134(636):1897–1910. https://doi.org/10.1002/qj.325

Kumar Das A, Bhowmick M, Kundu PK, Roy Bhowmik SK (2014) Verification of WRF rainfall forecasts over India during monsoon 2010: CRA method. Geofizika 31(2):106–126. https://doi.org/10.15233/gfz.2014.31.6

Lynch P (2008) The origins of computer weather prediction and climate modeling. J Comput Phys 227(7):3431–3444. https://doi.org/10.1016/j.jcp.2007.02.034

Madala S, Satyanarayana ANV, Rao TN (2014) Performance evaluation of PBL and cumulus parameterization schemes of WRF ARW model in simulating severe thunderstorm events over Gadanki MST radar facility—case study. Atmos Res 39:1–17. https://doi.org/10.1016/j.atmosres.2013.12.017

McCaul EW Jr, Goodman SJ, LaCasse KM, Cecil DJ (2009) Forecasting lightning threat using cloud-resolving model simulations. Weather Forecast 24(3):709–729. https://doi.org/10.1175/2008WAF2222152.1

Mielke Jr PW (1984) 34 Meteorological applications of permutation techniques based on distance functions. In: Krishnaiah PR, Sen PK (eds) Handbook of statistics. 4th vol. ScienceDirect, pp. 813–830. https://doi.org/10.1016/S0169-7161(84)04036-0

Mielke PW Jr (1991) The application of multivariate permutation methods based on distance functions in the earth sciences. Earth Sci Rev 31(1):55–71. https://doi.org/10.1016/0012-8252(91)90042-E

Mitra AK, Kaushik N, Singh AK, Parihar S, Bhan SC (2018) Evaluation of INSAT-3D satellite derived precipitation estimates for heavy rainfall events and its validation with gridded GPM (IMERG) rainfall dataset over the Indian region. Remote Sens Appl Soc Environ 9:91–99. https://doi.org/10.1016/j.rsase.2017.12.006

Mohan M, Bhati S (2011) Analysis of WRF model performance over subtropical region of Delhi, India. Adv Meteorol 2011:1–13. https://doi.org/10.1155/2011/621235

Naabil E, Lamptey BL, Arnault J, Olufayo A, Kunstmann H (2017) Water resources management using the WRF-hydro modelling system: case-study of the Tono dam in West Africa. J Hydrol Reg Stud 12:196–209. https://doi.org/10.1016/j.ejrh.2017.05.010

Nash JE, Sutcliffe JV (1970) River flow forecasting through conceptual models part I—a discussion of principles. J Hydrol 10(3):282–290. https://doi.org/10.1016/0022-1694(70)90255-6

Orrell D, Smith L, Barkmeijer J, Palmer TN (2001) Model error in weather forecasting. Nonlin Process Geophys 8(6):357–371. https://doi.org/10.5194/npg-8-357-2001

Paredes-Trejo FJ, Barbosa HA, Kumar TL (2017) Validating CHIRPS-based satellite precipitation estimates in Northeast Brazil. J Arid Environ 139:26–40. https://doi.org/10.1016/j.jaridenv.2016.12.009

Paul S, Wang CC, Chien FC, Lee DI (2018) An evaluation of the WRF Mei-yu rainfall forecasts in Taiwan, 2008–2010: differences in elevation and sub-regions. Meteorolog Appl 25(2):269–282. https://doi.org/10.1002/met.1689

Powers JG (2007) Numerical prediction of an Antarctic severe wind event with the weather research and forecasting (WRF) model. Mon Weather Rev 135(9):3134–3157. https://doi.org/10.1175/MWR3459.1

Powers JG, Klemp JB, Skamarock WC, Davis CA, Dudhia J, Gill DO, Coen JL, Gochis DJ, Ahmadov R, Peckham SE, Grell GA (2017) The weather research and forecasting model: overview, system efforts, and future directions. Bull Am Meteor Soc 98(8):1717–1737. https://doi.org/10.1175/BAMS-D-15-00308.1

Prakash S, Mitra AK, Pai DS, AghaKouchak A (2016) From TRMM to GPM: how well can heavy rainfall be detected from space? Adv Water Resour 88:1–7. https://doi.org/10.1016/j.advwatres.2015.11.008

Raju PVS, Potty J, Mohanty UC (2011) Sensitivity of physical parameterizations on prediction of tropical cyclone Nargis over the Bay of Bengal using WRF model. Meteorol Atmos Phys 113(3–4):125–137. https://doi.org/10.1007/s00703-011-0151-y

Richardson LF (1922) Weather prediction by numerical process. Cambridge University Press, Cambridge

Rife DL, Davis CA (2005) Verification of temporal variations in mesoscale numerical wind forecasts. Mon Weather Rev 133(11):3368–3381. https://doi.org/10.1175/MWR3052.1

Roberts NM, Lean HW (2008) Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. Mon Weather Rev 136(1):78–97. https://doi.org/10.1175/2007MWR2123.1

Shrestha DL, Robertson DE, Wang QJ, Pagano TC, Hapuarachchi HAP (2013) Evaluation of numerical weather prediction model precipitation forecasts for short-term streamflow forecasting purpose. Hydrol Earth Syst Sci 17(5):1913–1931. https://doi.org/10.5194/hess-17-1913-2013

Shuman FG (1989) History of numerical weather prediction at the National Meteorological Center. Weather Forecast 4(3):286–296. https://doi.org/10.1175/1520-0434(1989)004%3c0286:HONWPA%3e2.0.CO;2

Sikder S, Hossain F (2016) Assessment of the weather research and forecasting model generalized parameterization schemes for advancement of precipitation forecasting in monsoon-driven

river basins. J Adv Model Earth Syst 8(3):1210–1228. https://doi.org/10.1002/2016MS000678

Skamarock WC, Klemp JB, Dudhia J, Gill DO, Barker D, Duda MG et al (2008) A description of the advanced research WRF Version 3 (No. NCAR/TN-475+STR). University Corporation for Atmospheric Research. https://doi.org/10.5065/D68S4MVH. https://opensky.ucar.edu/islandora/object/technotes%3A500/datastream/PDF/view

Skofronick-Jackson G, Petersen WA, Berg W, Kidd C, Stocker EF, Kirschbaum DB, Kakar R, Braun SA, Huffman GJ, Iguchi T, Kirstetter PE (2017) The global precipitation measurement (GPM) mission for science and society. Bull Am Meteor Soc 98(8):1679–1695. https://doi.org/10.1175/BAMS-D-15-00306.1

Srinivas CV, Hariprasad D, Bhaskar Rao DV, Anjaneyulu Y, Baskaran R, Venkatraman B (2013) Simulation of the Indian summer monsoon regional climate using advanced research WRF model. Int J Climatol 33(5):1195–1210. https://doi.org/10.1002/joc.3505

Srivastava PK, Han D, Rico-Ramirez MA, O'Neill P, Islam T, Gupta M, Dai Q (2015) Performance evaluation of WRF-Noah Land surface model estimated soil moisture for hydrological application: synergistic evaluation using SMOS retrieved soil moisture. J Hydrol 529:200–212. https://doi.org/10.1016/j.jhydrol.2015.07.041

Tan J, Petersen WA, Tokay A (2016) A novel approach to identify sources of errors in IMERG for GPM ground validation. J Hydrometeorol 17(9):2477–2491. https://doi.org/10.1175/JHM-D-16-0079.1

Taylor KE (2001) Summarizing multiple aspects of model performance in a single diagram. J Geophys Res Atmos 106(D7):7183–7192. https://doi.org/10.1029/2000JD900719

Tie X, Brasseur G, Ying Z (2010) Impact of model resolution on chemical ozone formation in Mexico City: application of the WRF-Chem model. Atmos Chem Phys 10(18):8983–8995. https://doi.org/10.5194/acp-10-8983-2010

Tolstykh MA, Frolov AV (2005) Some current problems in numerical weather prediction. Izvest Atmos Ocean Phys 41(3):285–295

Vasić S, Lin CA, Zawadzki I, Bousquet O, Chaumont D (2007) Evaluation of precipitation from numerical weather prediction models and satellites using values retrieved from radars. Mon Weather Rev 135(11):3750–3766. https://doi.org/10.1175/2007MWR1955.1

Wang H, Mu M, Huang XY (2011) Application of conditional non-linear optimal perturbations to tropical cyclone adaptive observation using the weather research forecasting (WRF) model. Tellus A Dyn Meteorol Oceanogr 63(5):939–957. https://doi.org/10.1111/j.1600-0870.2011.00536.x

Wang W, Lu H, Zhao T, Jiang L, Shi J (2017) Evaluation and comparison of daily rainfall from latest GPM and TRMM products over the Mekong River Basin. IEEE J Sel Topics Appl Earth Observ Remote Sens 10(6):2540–2549. https://doi.org/10.1109/JSTARS.2017.2672786

Watterson IG (1996) Non-dimensional measures of climate model performance. Int J Climatol 16(4):379–391. https://doi.org/10.1002/(sici)1097-0088(199604)16:4%3c379::aid-joc18%3e3.0.co;2-u

Wernli H, Paulat M, Hagen M, Frei C (2008) SAL—a novel quality measure for the verification of quantitative precipitation forecasts. Mon Weather Rev 136(11):4470–4487. https://doi.org/10.1175/2008MWR2415.1

Wilks DS (2011) Statistical methods in the atmospheric sciences, vol 100. Academic press, USA

Willmott CJ (1981) On the validation of models. Phys Geogr 2(2):184–194. https://doi.org/10.1080/02723646.1981.10642213

Willmott CJ, Matsuura K (2005) Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate Res 30:79–82. https://doi.org/10.3354/cr030079

Willmott CJ, Matsuura K, Robeson SM (2009) Ambiguities inherent in sums-of-squares-based error statistics. Atmos Environ 43(3):749–752. https://doi.org/10.1016/j.atmosenv.2008.10.005

Willmott CJ, Robeson SM, Matsuura K (2012) A refined index of model performance. Int J Climatol 32(13):2088–2094. https://doi.org/10.1002/joc.2419

Zhao Q, Liu Z, Ye B, Qin Y, Wei Z, Fang S (2009) A snowmelt runoff forecasting model coupling WRF and DHSVM. Hydrol Earth Syst Sci 13(10):1897–1906. https://doi.org/10.5194/hess-13-1897-2009

Zhou G, Xu J, Xie Y, Chang L, Gao W, Gu Y, Zhou J (2017) Numerical air quality forecasting over eastern China: an operational application of WRF-Chem. Atmos Environ 153:94–108. https://doi.org/10.1016/j.atmosenv.2017.01.020