BCWC RESEARCH REPORT

BCWC/RR/01/2024

Verification of probabilistic precipitation forecast by the NCMRWF global ensemble prediction system over Bhutan

Sukanya Bose, Gauri Shanker, Monju Subba, M. Venktarami Reddy, Abhijit Sarkar, Raghavendra Ashrit, V. S. Prasad

July 2024

Bangladesh Bhutan India Myanmar Nepal Sri Lanka Thailand

BIMSTEC Centre for Weather and Climate (BCWC)

National Centre for Medium Range Weather Forecasting,
Ministry of Earth Sciences, Government of India
A-50, Sector-62, NOIDA-201 309, INDIA

Verification of probabilistic precipitation forecast by the NCMRWF global ensemble prediction system over Bhutan

Sukanya Bose[1], Gauri Shanker[2,3], Monju Subba[4], M. Venktarami Reddy[2], Abhijit Sarkar[2], Raghavendra Ashrit[2], V. S. Prasad[2]

[1]Department of Atmospheric Sciences, University of Calcutta, 51/2 Hazra Road, Kolkata-700 019

[2]National Centre for Medium Range Weather Forecasting (NCMRWF), (Ministry of Earth Sciences, Government of India), A-50, Sector-62, Noida 201 309, India

[3]Department of Geophysics, Banaras Hindu University, Varanasi 221 005, Uttar Pradesh, India

[4]Weather and Climate Services Division, National Centre for Hydrology and Meteorology, Thimphu, Bhutan

July 2024

BIMSTEC Centre for Weather and Climate (BCWC)

National Centre for Medium Range Weather Forecasting

Ministry of Earth Sciences, Government of India

A-50, Sector-62, Noida-201 309, INDIA

Data Control Sheet

1	Name of the Institute	National Center for Medium Range Weather Forecasting
2	Document Number	BCWC/RR/01/2024
3	Date of Publication	July 2024
4	Title of the document	Verification of probabilistic precipitation forecast by the NCMRWF global ensemble prediction system over Bhutan
5	Type of the document	BCWC Research Report
6	Number of pages, figures and Tables	33 pages, 11 figures, 1 Table
7	Authors	Sukanya Bose, Gauri Shanker, Monju Subba, M. Venktarami Reddy, Abhijit Sarkar, Raghavendra Ashrit, V. S. Prasad
8	Originating Unit	BIMSTEC Centre for Weather and Climate (BCWC), National Centre for Medium Range Weather Forecasting (NCMRWF), A-50, Sector-62, NOIDA 201 309, India
9	Abstract	Bhutan, a South Asian country located in the Eastern Himalayas, is an important member of the Bay of Bengal Initiative of Multi–sectorial Technical and Economic Cooperation (BIMSTEC). BIMSTEC Centre for Weather and Climate (BCWC), established at National Centre for Medium Range Weather Forecasting (NCMRWF), India promotes collaborative research in the field of weather prediction and climate change among the scientists of BIMSTEC member countries. Through the present work BCWC brings together the researchers of Bhutan and India to verify the probabilistic precipitation forecast by NCMRWF Global Ensemble Prediction System (NEPS-G) for the 2020 Monsoon season over Bhutan. The Global Precipitation Measurement (GPM) IMERG Final Run data and the satellite-gauge merged gridded dataset prepared at NCMRWF have been used as verifying observations. The gauge data have been provided by National Centre for Hydrology and Meteorology (NCHM), Bhutan. Various standard probabilistic forecast verification metrics like RMSE-Spread relationship, Brier Skill Score (BSS), Rank Histogram, Reliability Diagram and Relative Operating Characteristic (ROC) curve have been used to carry out the verification. Since the verifying data over the mountainous North Bhutan region is not reliable the validation work is mainly confined to South Bhutan region. BSS of the model is found to be positive nearly at all forecast lead times for higher thresholds (>25 mm/day and >45 mm/day) values of precipitation. The ensembles are under-dispersive but the forecasts are generally reliable. The ROC curves show that the forecasts have good skill in discriminating events from non-events, particularly for heavier precipitation thresholds. The ROC score of the probabilistic forecast is found to be higher than the control forecast.
10	References	22
11	Security classification	Unrestricted
12	Distribution	General

Table of Contents

Abstract

Bhutan, a South Asian country located in the Eastern Himalayas, is an important member of the Bay of Bengal Initiative of Multi–sectorial Technical and Economic Cooperation (BIMSTEC). BIMSTEC Centre for Weather and Climate (BCWC), established at National Centre for Medium Range Weather Forecasting (NCMRWF), India promotes collaborative research in the field of weather prediction and climate change among the scientists of BIMSTEC member countries. Through the present work BCWC brings together the researchers of Bhutan and India to verify the probabilistic precipitation forecast by NCMRWF Global Ensemble Prediction System (NEPS-G) for the 2020 Monsoon season over Bhutan. The Global Precipitation Measurement (GPM) IMERG Final Run data and the satellite-gauge merged gridded dataset prepared at NCMRWF have been used as verifying observations. The gauge data have been provided by National Centre for Hydrology and Meteorology (NCHM), Bhutan. Various standard probabilistic forecast verification metrics like RMSE-Spread relationship, Brier Skill Score (BSS), Rank Histogram, Reliability Diagram and Relative Operating Characteristic (ROC) curve have been used to carry out the verification. Since the verifying data over the mountainous North Bhutan region is not reliable the validation work is mainly confined to South Bhutan region. BSS of the model is found to be positive nearly at all forecast lead times for higher thresholds (>25 mm/day and >45 mm/day) values of precipitation. The ensembles are under-dispersive but the forecasts are generally reliable. The ROC curves show that the forecasts have good skill in discriminating events from non-events, particularly for heavier precipitation thresholds. The ROC score of the probabilistic forecast is found to be higher than the control forecast (i.e., NCMRWF's operational global deterministic model (NCUM-G) forecast).

सारांश

भूटान, पूर्वी हिमालय में स्थित एक दक्षिण एशियाई देश, बंगाल की खाड़ी बहु-क्षेत्रीय तकनीकी और आर्थिक सहयोग पहल (बिम्सटेक) का एक महत्वपूर्ण सदस्य है। भारत के राष्ट्रीय मध्यम अवधि मौसम पूर्वानुमान केंद्र (एनसीएमआरडब्ल्यूएफ) में स्थापित बिम्सटेक मौसम और जलवायु केंद्र (बीसीडब्ल्यूसी), बिम्सटेक सदस्य देशों के वैज्ञानिकों के बीच मौसम के पूर्वानुमान और जलवायु परिवर्तन के क्षेत्र में सहयोगात्मक अनुसंधान को बढ़ावा देता है। वर्तमान कार्य के माध्यम से बीसीडब्ल्यूसी भूटान में 2020 के मानसून सीज़न के लिए एनसीएमआरडब्ल्यूएफ ग्लोबल एन्सेम्बल प्रेडिक्शन सिस्टम (एनईपीएस-जी) द्वारा संभावित वर्षा पूर्वानुमान को सत्यापित करने के लिए भूटान और भारत के शोधकर्ताओं को एक साथ लाता है। वैश्विक वर्षा मापन (जीपीएम) आईएमईआरजी फाइनल रन डेटा और एनसीएमआरडब्ल्यूएफ में तैयार सैटेलाइट-गेज मर्ज किए गए ग्रिडेड डेटासेट का उपयोग अवलोकन के रूप में किया गया है पूर्वानुमान को सत्यापित करने के लिए। गेज डेटा, राष्ट्रीय जल विज्ञान और मौसम विज्ञान केंद्र (एनसीएचएम), भूटान द्वारा प्रदान किया गया है। सत्यापन करने के लिए आरएमएसई-स्प्रेड रिलेशनशिप, ब्रियर स्किल स्कोर (बीएसएस), रैंक हिस्टोग्राम, विश्वसनीयता आरेख और सापेक्ष परिचालन विशेषता (आरओसी) वक्र जैसे विभिन्न मानक संभाव्य पूर्वानुमान सत्यापन मेट्रिक्स का उपयोग किया गया है। चूंकि पर्वतीय उत्तरी भूटान क्षेत्र पर सत्यापन डेटा विश्वसनीय नहीं है, इसलिए सत्यापन कार्य मुख्य रूप से दक्षिण भूटान क्षेत्र तक ही सीमित है। मॉडल का बीएसएस वर्षा की उच्च सीमा (>25 मिमी/दिन और >45 मिमी/दिन) मूल्यों के लिए लगभग सभी पूर्वानुमानित लीड समय पर सकारात्मक पाया गया है। एन्सेम्बल अल्प-विस्तारित हैं लेकिन पूर्वानुमान आम तौर पर विश्वसनीय होते हैं। आरओसी वक्र दिखाते हैं कि पूर्वानुमानों में घटनाओं को गैर-घटनाओं से अलग करने में अच्छा कौशल है, खासकर भारी वर्षा सीमा के लिए। संभाव्य पूर्वानुमान का आरओसी स्कोर नियंत्रण पूर्वानुमान (यानी, एनसीएमआरडब्ल्यूएफ के परिचालन वैश्विक नियतात्मक मॉडल (एनसीयूएम-जी) पूर्वानुमान) से अधिक पाया गया है।

1. Introduction

The landscape of Bhutan extends from 26.64^0 N to 28.2^0 N by latitudes and 88.74^0 E to 92.16^0 E by longitudes and it lies in the Northern Temperate Zone. Located in the Eastern Himalayas in Southeast Asia, Bhutan experiences a diverse climate due to its altitude and topography. A very pleasant weather generally persists throughout the elevation of the valley except in the northern and southern extremes. The middle valley of Bhutan experiences a pleasant temperate climate with moderate changes between winter and summer. A map of Bhutan is presented in Figure 1.

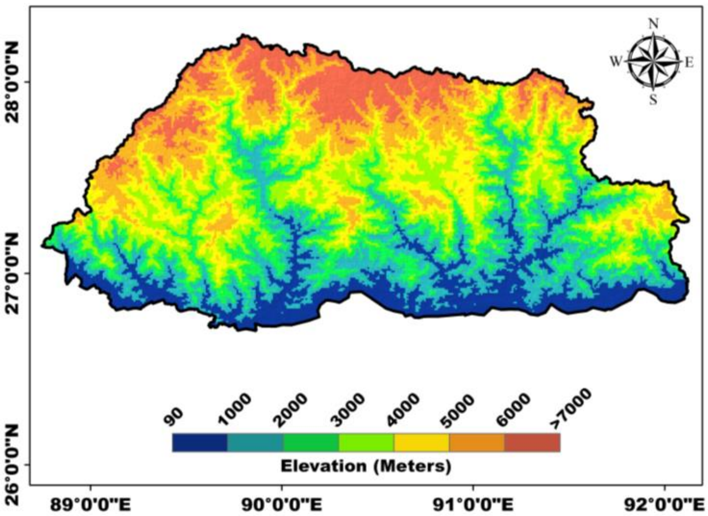


Figure 1. *Map of Bhutan.*

Bhutan has an extremely varied climate due to the vast difference in altitude and also due to the impact of the North Indian Monsoon. In the northern part of Bhutan where the mountains rise up to 7000m, the climate is similar to that of arctic and has alpine and subalpine climates. In the southern part, the country experiences a hot and humid summer and cool winter, with usual heavy monsoon rains. The south of Bhutan experiences a tropical climate during the monsoons. The Himalayas create a barrier effect for the south-westerly monsoon winds coming from the Bay of Bengal, resulting in orographic lifting which is responsible for the heavy rainfall in the southern foothills of Bhutan and a rain-shadow zone in the North.

The Bay of Bengal Initiative of Multi–sectorial Technical and Economic Cooperation (BIMSTEC) was established on June 6, 1997. It is a regional organization that comprises seven member states namely, Bangladesh, India, Sri Lanka, Thailand, Myanmar, Bhutan and Nepal. There are several sectors in which BIMSTEC perform, and for every such sector, there are several centres set up in different parts of the member countries. “Security” is one such sector and it includes a subsector “Disaster Management”. The BIMSTEC Centre for Weather and Climate (BCWC) located at the National Centre for Medium Range Weather Forecasting (NCMRWF), Noida, Uttar Pradesh, India has been established under the subsector “Disaster Management”. BCWC aims at the following objectives:

- To Facilitate and foster cooperation amongst the BIMSTEC members to work more competently in fields of both fundamental and scientific research related to weather predictions and climate change, for mutual benefit.
- To Promote an enriched spectrum of research pertaining to weather and climate.
- To Publish the important research outcomes under the framework of BIMSTEC, on weather and climate.

Bhutan has joined the BIMSTEC in February 2004. Since then, it has become a very important member of the organization. Bhutan collaborates with other member countries in various sectors to address common challenges, promote regional cooperation and facilitation of trade and most importantly promote economic growth.

A high resolution (~ 12 km) global ensemble prediction system, NEPS-G is run operationally at NCMRWF on the high-performance computing system “Mihir” to generate medium range (for forecast lead time of 10 days) probabilistic forecast of the weather. A brief description of this ensemble prediction system is available in section 2. The performance of NEPS-G over the Indian subcontinent for different seasons and events has been studied before by a few investigators (Mamgain et al., 2020a; Chakraborty et al., 2021; Shanker et al., 2022). The present study aims to study the performance of NEPS-G in forecasting precipitation over Bhutan for the Monsoon season (June, July, August and September) of 2020. A number of probabilistic verification metrics have been used to carry out the forecast validation. The details of the methodology and the verifying data are provided in section 2. It is the diverse topography and climatology of Bhutan that drew our attention to do a specific study of the forecasts made over the years. Moreover, the authors are not aware of any study that includes verification of probabilistic precipitation forecast over Bhutan by a global ensemble forecasting system.

2. Model, Data and Methodology

The global ensemble prediction system of NCMRWF (NEPS-G) is based on the global version of Met Office Global and Regional Ensemble Prediction System (MOGREPS) of Met Office, UK. The horizontal resolution of NEPS-G is about 12 km and it has 23 (1 control +22 perturbed) ensemble members. The operational deterministic model (NCUM) forecast is used as the control forecast of NEPS-G. The perturbations to the initial condition for the 22 perturbed ensemble members are generated by the Ensemble Transform Kalman Filtering (ETKF) method (Bishop et al., 2001). These perturbations are added to the initial condition generated by the Hybrid 4D VAR data assimilation method to prepare 22 perturbed initial conditions at 00, 06, 12 and 18 UTC every day. The control member (NCUM) runs from the unperturbed initial condition (IC) generated by the Hybrid 4D VAR data assimilation method. Out of these 22 perturbed members, 11 members are integrated forward in time for 10 days forecast lead time from the initial conditions of 00 UTC and 12 UTC every day. Eleven perturbed members and one control member running from 00 UTC initial condition of current day combine with the eleven perturbed members running from 12 UTC initial condition of previous day to form a 23-member ensemble forecast. The Stochastic Kinetic Energy Back Scattering (SKEB) (Tennant et al., 2011)and Stochastic perturbation of Physics Tendency (SPT) (Sanchez et al.,

2016) schemes are used to represent the uncertainty in model physics. A more detailed description of NEPS-G is available at Mamgain et al. (2020b).

NEPS-G provides 23-member ensemble forecast of 3-hourly accumulated precipitation up to 10 days forecast lead time. From these forecast data, 24-hourly accumulated daily precipitation for each forecast lead day has been prepared for the four monsoon months (JJAS) of the year 2020. The Global Precipitation Measurement (GPM) is a joint mission between the National Aeronautics and Space Administration (NASA) and the Japan Aerospace Exploration Agency (JAXA), providing the next-generation global observations of rain, snow and other precipitation data (Hou et al., 2014). The Integrated Multi-satellite Retrievals for GPM (IMERG) is a multi-satellite precipitation product which is obtained by combining passive microwave (PMW) and infrared (IR) data of GPM constellation satellites. IMERG data is available at a spatial resolution of 0.1° and a temporal resolution of 30 minutes. The details of data and algorithm description can be found in (Huffman et al., 2020).

The GPM precipitation data has been classified into Levels 1, 2 and 3. We have taken into account the Level 3 data for our verification. Level 3 data is just a spatial and temporal resampling of the geophysical parameters of level 1 and 2 data.

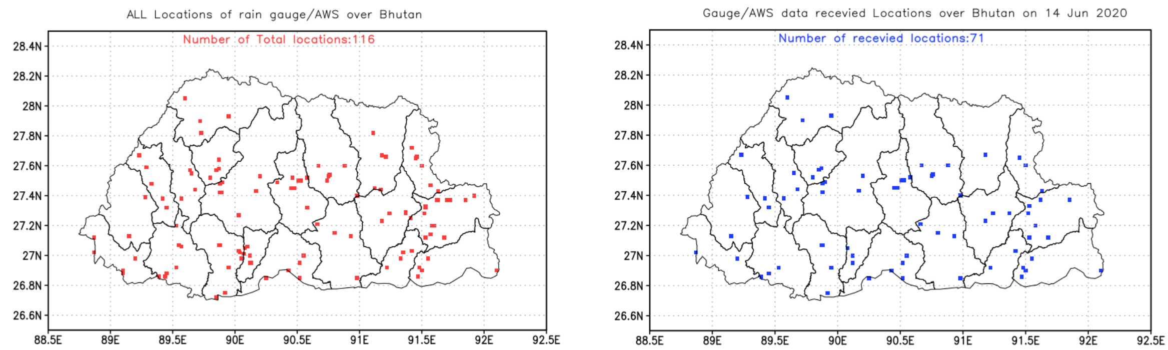
The level 3 data has been further classified into 3 types (Huffman et al., 2020) –

1. The IMERG Early Run – this gives real-time low latency gridded global multi-satellite precipitation estimates
2. The IMERG Late Run – this gives near real–time gridded global multi–satellite precipitation estimates with quasi-Lagrangian time interpolation.
3. The IMERG Final run – this is a research-quality gridded global multi–satellite precipitation estimates with quasi–Lagrangian time interpolation, gauge data and climatological adjustment.

For the purpose of our research on the verification of NEPS-G precipitation forecast over Bhutan, we have used the IMERG Final Run data as the verifying observation. The verifying IMERG data will be referred as 'GPM' data hereafter. The data have been directly downloaded from the official site of NASA Global Precipitation Measurement Data Directory. The link to the site has been cited below – https://gpm.nasa.gov/data/directory

Daily gauge precipitation data are collected at different stations by the National Centre for Hydrology and Meteorology (NCHM), Bhutan. A satellite (GPM) – gauge (shared by NCHM, Bhutan) merged gridded dataset of about 10 km resolution has been prepared at NCMRWF (Mitra et al., 2009). This gridded data has been up-scaled to the model resolution (12 km) and also used as verifying observation. A brief description of the methodology used to create NCMRWF merged Satellite and Gauge precipitation for Bhutan (NMSG-B) is given below.

The first guess used in the successive correction method for rainfall analysis is taken from the GPM satellite. In the present work, IMERG version 6 data product is used for NMSG-B. The other data used are the 24-h accumulated rainfall values from rain gauge/AWS observations over Bhutan. The daily rainfall records of 116 rain gauge stations with varying availability periods were used for the present study. However, the data density varies from day to day during the period JJAS 2020. These 116 stations include both the hydro-meteorology observatories and Agromet observatories. The locations of all the 116 rain gauge stations considered for preparing the new merged rainfall data are shown in Figure 2(a). As seen in Figure 2(a), the spatial density of the station points is not uniform throughout the country. The density of the stations is relatively high in south Bhutan and low over northern Bhutan of the country. Figure 2(b) shows the geographical distribution of such gauge observations on a typical day (14th June 2020).

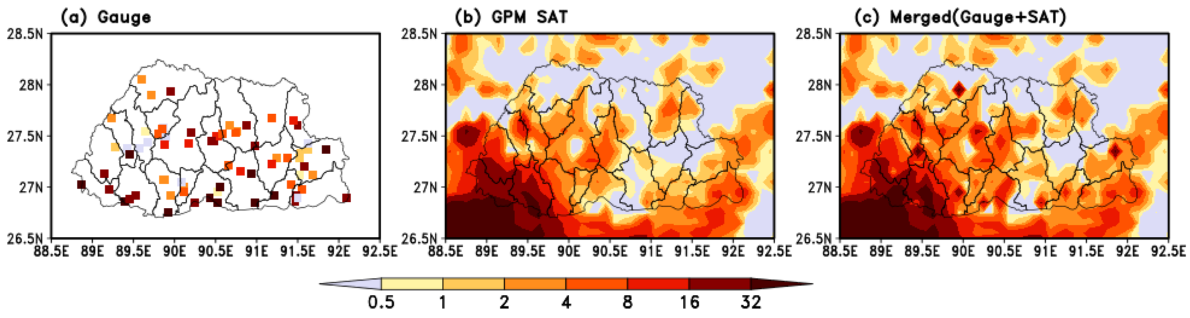


(a) (b)

Figure 2. *(a) Total Network of rain gauge/AWS stations over Bhutan and (b) Network of 71 rain gauge/AWS stations used for the development of NMSG-B on 14th Jun 2020*

The objective analysis technique for rainfall used (Tripoli & Krishnamurti, 1975; Krishnamurti et al. 1983; Mitra et al., 2003;) in the present study is based on the successive correction method of Cressman (1959) which modifies the initial guess (satellite estimates) based on the observations (rain gauge). In this method, the first guess value for each station is obtained by interpolating the satellite measurements. The difference between the observed value and the first guess provides the error estimate at the station location. The corrections at the grid points are obtained by using the successive iterative corrections on these error estimates. Mitra et al., (2003, 2009) have discussed the details of weights and interpolations for this successive iteration correction.

The spatial distributions of mean daily rainfall from gauge, GPM satellite and NMSG-B are presented for 14th Jun 2020 over Bhutan (Figure 3). Figure 3 indicates that a few stations in the country had more than 16 mm/day of rainfall. Compared to the gauge station analysis (Figure 3(a)), the GPM analysis (Figure 3 (b)) underestimates the rainfall over most of the country. In the NMSG-B analysis, areas with rainfall amounts greater than 16 mm/day match well with the Gauge analysis and are close to GPM. Further, the mean rainfall averaged over the whole Bhutan region for 14th June 2020 from gauge-based, GPM satellite and NMSG-B are 10.7 mm, 4.4 mm and 7.2 mm, respectively (Figure 4). This indicates that after merging the satellite and gauge data (NMSG-B) there is an improvement in verifying the dataset compared to GPM satellite data. Reddy et al., (2019, 2022) have mentioned that the satellite and gauge merged (NMSG) product estimated the rainfall better than the other satellite products over India. Hereafter, NMSG-B data will be referred to as 'Merged' data.



(a) (b) (c)

Figure 3. *Spatial distribution of rainfall (mm/day) from (a) surface Gauge/AWS (b) GPM Satellite and (c) NMSG-B over Bhutan on 14th Jun 2020.*

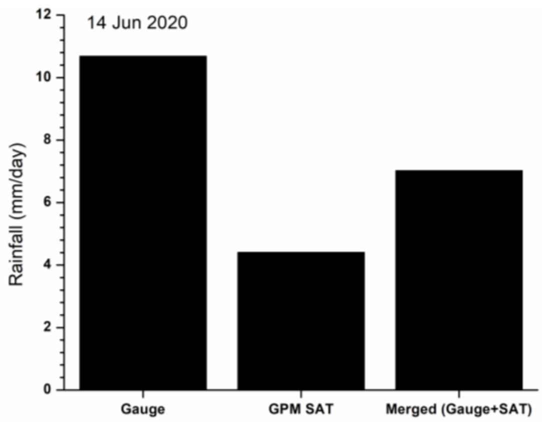


Figure 4. *Mean rainfall (mm day^{-1}) averaged over the whole Bhutan region from (a) surface Gauge/AWS (b) GPM Satellite and (c) NMSG-B on 14th Jun 2020.*

The forecast characteristic of a forecasting system cannot be fully understood from the computation of a single verification metric. So various standard probabilistic verification metrics have been used to validate the precipitation forecast of NEPS-G over Bhutan. These metrics are: (1) Root Mean Squared Error (RMSE) - Spread relationship, (2) Brier Skill Score

(BSS), (3) Rank Histogram, (4) Reliability Diagram, (5) Relative Operating Characteristics (ROC) and Area Under the ROC Curve (ROC Score).

3. Results and Discussion

3.1. Relationship between RMSE of ensemble mean and ensemble spread

Root mean squared error (RMSE) or the square root of mean square error is the measure of the amplitude of the error associated with the forecast which tells us about the accuracy of the forecast. This does not suggest the direction of the forecast. The root mean square error is given by:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(F_i - O_i)^2} \tag{1}$$

where N is the total number of forecast and observation pairs. F_i is the ensemble mean forecast and O_i is the observation data for that particular grid.

The forecast variance at i^{th} grid point is given by :

$$\sigma_i{}^2 = \frac{1}{n}\{\sum_k (F_i - F_k)^2\} \tag{2}$$

The mean forecast spread is given by

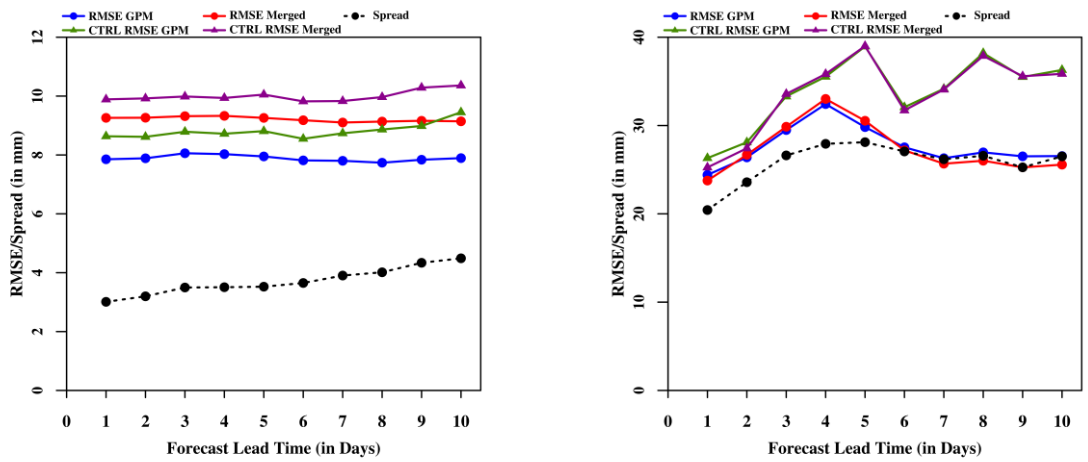
$$S = \left[\frac{1}{N}\sum_i \sigma_i^2\right]^{\frac{1}{2}} \tag{3}$$

where n is the number of ensemble members or ensemble size, F_i is the ensemble mean forecast at i^{th} grid point and F_k is the forecast of k^{th} ensemble member. According to Palmer (2019), the mean spread should be equal to the root mean square error of the ensemble mean for a perfect ensemble prediction. The ratio of spread to RMSE indicates overconfidence or underconfidence if it is <1 or > 1, respectively.

A smaller value of the root mean square error indicates more accuracy of the forecast. Since no ensemble prediction system can take into consideration all the uncertainties associated with

NWP modelling the ensemble forecasts are generally under-dispersive and the spread is smaller than RMSE. Since the atmosphere and its model are chaotic systems both RMSE and spread increase with forecast lead time.

Figure 5 shows the variation of RMSE and Spread with forecast lead time for precipitation over North (Figure 5(a)) and South Figure 5(b)) Bhutans using both GPM and Merged Data as verifying observations. The RMSE computed between the NEPS-G ensemble mean and GPM data is represented by RMSE GPM, while the RMSE computed between the NEPS-G ensemble mean and Merged data is represented by RMSE Merged. The RMSE computed between the control forecast and GPM data is represented by CTRL RMSE GPM, whereas the RMSE computed between the control forecast and the satellite-gauge merged data is represented by CTRL RMSE Merged. Spread is a representation of the NEPS-G ensemble spread.



(a) (b)

***Figure 5**. Variation of RMSE and Spread with forecast lead time for precipitation over (a) North Bhutan and (b) South Bhutan using GPM and Merged data as verifying observation.*

For North Bhutan (Figure 5(a)), we see that the RMSE GPM shows the lowest RMSE values while CTRL RMSE Merged shows the highest. This indicates that the control forecast shows a large error when validated against the Merged dataset. The dashed line representing the ensemble spread always lies below the RMSE lines indicating the underdispersion of the member forecasts. The RMSE and spread should increase with forecast lead time, but in the case of North Bhutan although spread shows an increasing tendency with forecast lead time the RMSE of both control and ensemble mean remain nearly constant. One of the possible reasons for such a constant RMSE may be a small sample size; another can be the variable itself, i.e., accumulated precipitation. In previous study by Shanker et. al. 2022, it was found that the rate of growth of error/spread of NEPS-G rainfall forecast over the Indian region is quite low when compared to other variables such as u and v wind (at 200 and 850 hPa) over the Indian region.

For South Bhutan, the RMSE of control forecast shows a general increasing trend with forecast lead time. The RMSE of ensemble mean forecast (RMSE GPM and RMSE Merged) initially increases with forecast lead time till day four. Then it starts decreasing and remains constant. For control forecast (CTRL RMSE GPM and CTRL RMSE Merged). RMSE values increase till day 5 then decrease on day 6 then again increase after that. The sharp variations of the RMSE values of both ensemble mean and control forecasts over a short period of time may be due to the small sample size (Atger, 2004). The smaller RMSE value of ensemble mean forecast compared to the control forecast at all forecast lead times indicates the superiority of ensemble mean forecast over control. Although the forecast is underdispersive with respect to ensemble mean forecast during first few forecast lead days both the RMSE curves (RMSE GPM and RMSE Merged) show same tendency of variation with forecast lead time.

3.2. Brier Score (BS) and Brier Skill Score (BSS)

The brier score is used to measure the accuracy of the probabilistic forecast of a 'yes' or 'no' event. It is defined as:

$$BS = \frac{1}{N}\sum_{i=1}^{N}(p_i - o_i)^2 \tag{4}$$

where, N is the number of 'forecast – observation' pairs, p_i is the forecast probability of occurrence of an event, o_i is the observed outcome which is noted in binary, 1 if the event occurred and 0 if it did not. Thus, BS yields a value between 0 and 1. Brier Score is negatively oriented, i.e. smaller the BS, more will be the accuracy of the forecast. For a perfect forecast, BS is 0. As the skill of the forecast decreases with lead time, BS increases. The Brier score strongly depends on the climatological frequency of an event over an area. For events that have rare chances of occurrence, BS will show a low value but that does not necessarily mean that the forecast is skilful. The low value of BS in case of rare events is due to the large number of zero values of forecast-observation pairs.

Although we have not included the BS outcomes of the precipitation forecast during JJAS, 2020 over Bhutan, in this report, it was calculated for the primary verification and the calculation of Brier skill score (BSS). BSS is computed from the BS. It measures whether the forecast is more skilful than a reference forecast. It does so by comparing the BS of a forecasting system to that of a reference forecast system, which is typically the climatology. Unlike BS, the BSS is positively oriented. For a better improved forecast system, the score can have a maximum value of 1. It is the usual practice, to consider the climatological forecast as the reference. However, the skill of a new forecasting system can also be determined by calculating the score relative to the old forecast system (Wilks, 2019). The BSS is defined as

$$BSS = 1 - (BS / BS_{ref}) \quad (5)$$

where BS is the Brier score of the present forecasting system and BS_{ref} is the Brier score of the reference forecasting system.

The BSS for South Bhutan (Figure 6(a)) for the exceedance of 2.5mm/day of rainfall ranges between -0.2 and -0.27 for Merged data and -0.32 and -0.45 for GPM. The score decreases with forecast lead time. For the exceedance of 25mm/day of precipitation (Figure 6(b)), BSS calculated from GPM data starts with 0.15 on Day 1, shows a slight increase in score to 0.175 on Day 2, then again gradually decreases to 0.03 on Day 10 whereas BSS values

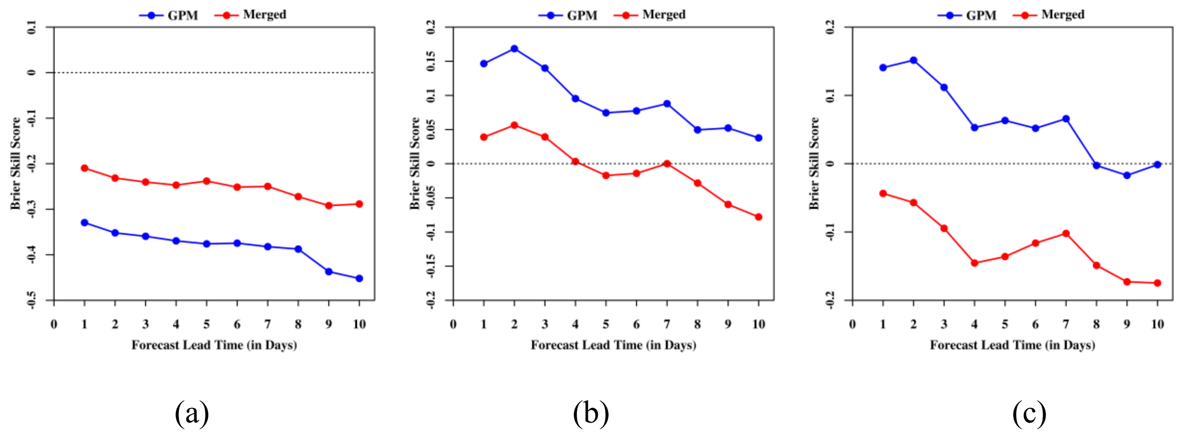


Figure 6. *Variation of Brier skill score with Forecast lead time for South Bhutan using GPM (blue) and Merged (red) data as verifying observations for precipitation exceeding (a) 2.5 mm/day, (b) 25 mm/day and (c) 45 mm/day.*

obtained from Merged data is negative from day 5 onwards. For precipitation exceeding 45mm/day (Figure 6(c)), the skill measured against GPM gradually reduces from about 0.15 on day-1 forecast to about 0.07 on day 7 forecast (day 8-10 being negative) whereas, in the case of Merged dataset, BSS values are negative at all lead times. The negative value of BSS for light precipitation (>2.5 mm/day) may be due to the positive bias of model over the region. For heavier rainfall categories better skill with respect to GPM may be because both model and GPM (as already mentioned in Section 1) under-estimates heavy precipitation amount.

The skill of the model is compromised to some extent in the case of our study over the whole Bhutan (results not shown). This is justified by the part-wise study of North Bhutan and South Bhutan. The verifying data over mountainous North Bhutan is not very reliable because GPM data under-estimates precipitation and density of gauge observations over North Bhutan is very low. Also, the numerical models show lower prediction skill over irregular topography of hilly region. Again, NEPS-G has a wet bias for low thresholds of precipitation. The sample size readily affects the verification of accuracy and reliability of the model forecast, hence the poor result over North Bhutan can be understood, which in turn had affected the forecasts over all of Bhutan on an average. So, we have presented the verification result of BSS only for South

Bhutan. In the next sections also, we will study the verification of forecasts only over South Bhutan.

3.3. Rank Histogram

Rank histogram or Talagrand histogram (Hamill, 2001) is used to determine where does the degree of observed intensity lies within the forecast distribution. For a specified domain, we gather the forecasted value of the parameter given by each ensemble member and then it is compared to the observation. Now, these observed and forecasted values are arranged in ascending order, the rank that the observation falls in is then plotted on a histogram graph. A rank histogram can be used to assess the consistency of ensemble forecasting system. It also describes the ensemble spread and bias. Depending on the shape of the rank histogram curve, we can say if the forecast is perfect or not.

For a perfect forecast the shape of rank histogram should be flat. A 'u' shaped rank histogram indicates under-dispersive and an 'n' shaped rank histogram indicates over-dispersive ensemble members. Taller bars at lower ranks indicate positive bias and at higher ranks indicate negative bias of the forecast. After plotting the rank histogram, we can test the flatness of the rank histogram using Pearson's chi-squared (χ^2) test (Wilks, 2019; Bröcker, 2018). Larger the value of χ^2 less uniform will be the histogram.

The Rank Histograms in Figure 7 for South Bhutan, for both GPM and Merged data as verifying observations, show a very sharp negative slope from which we can clearly note that the model had over forecasted the precipitation events on most of the cases. The u-shaped rank histograms also indicate that the ensemble members are under dispersive and hence the forecast is over-confident. For GPM and Merged data, the χ^2 values across South Bhutan are 75,440.55 and 51,005.38, respectively. Thus, when NEPS-G is validated against the Merged dataset, a lower value of χ^2 is obtained for South Bhutan, indicating a more uniform rank histogram.

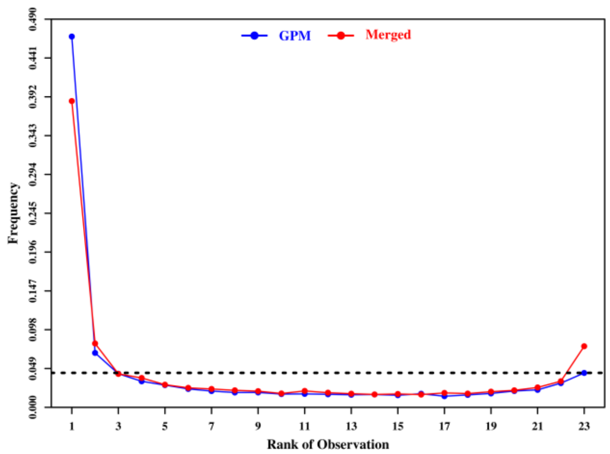


Figure 7. *The Rank Histogram curve for the precipitation over South Bhutan using GPM (blue) and Merged (red) data for Day 5 forecast lead time. A perfectly calibrated 22-member ensemble would have a flat histogram at 4.35% (dotted black line).*

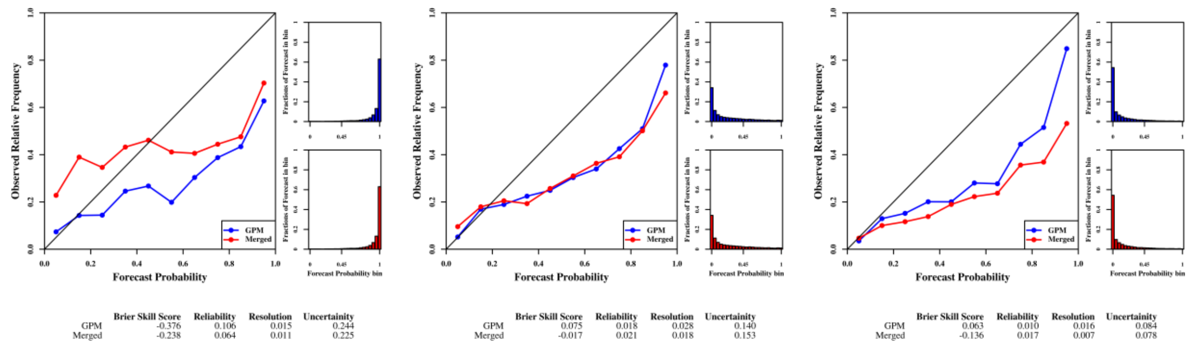
3.4. Reliability Diagram

The reliability diagram (Wilks, 2019) is an important metric used for the verification of forecast probabilities. It is conditioned on the forecast and tells us - given the forecast probability what is the relative observed frequency of the event. Reliability gives a check on how trustworthy the forecast is. If it says that the probability of occurrence of a certain event is x% on ‘n’ number of occasions and the actual occurrence of the event is found to be x% in those ‘n’ occasions and it holds true for all possible values of x, the model is considered to be perfectly reliable. This distribution of forecast and observation probability is portrayed by a graph called the reliability diagram. In a reliability diagram, also known as an attribute diagram, observed relative frequency is plotted against forecast probability.

The line of perfect reliability is the diagonal with slope 1. In such cases, the forecast probability exactly matches the relative observed frequency. If the reliability curve lies below the line of slope 1, or the line of perfect reliability, then the system has over-forecasted the events and if the reliability curve lies above the line of perfect reliability, the system has under-forecasted the events. A reliability diagram often comes with a sharpness diagram as a subplot. The sharpness diagram shows the number of times each probability value was predicted. For a value of 0 or 100%, it indicates that the forecast is very sharp.

The reliability diagram can be linked with the Rank histogram as well. The underconfident or overconfident forecast can be detected from the spread of the rank histogram. It can even tell us about the accuracy of the forecast from the Brier Score. The decomposition of the Brier score gives us the values of reliability, resolution and uncertainty. If the value of reliability is more, the Brier score will be more, which indicates lesser accuracy.

The Reliability Diagrams for South Bhutan using GPM and Merged data for precipitation exceeding 2.5 mm/day, 25 mm/day and 45 mm/day for day-5 forecast are shown in Figure 8(a), 8(b) and 8(c) respectively. The rainfall thresholds 2.5, 25 and 45 mm/day represent light, moderate and heavy precipitations. We have selected these values to study the model performance for precipitations of varying intensities. The reliability diagram was analysed for all 10-day lead times but for brevity, only day-5 forecast is presented here. In each figure, values of BSS, reliability terms, resolution terms and uncertainty terms are also given at the bottom for both GPM and Merged data as verifying observation.



(a) (b) (c)

Figure 8. *The reliability diagrams for South Bhutan using GPM (blue) and Merged (red) data for precipitation exceeding (a) 2.5mm/day (b) 25mm/day and (c) 45 mm/day for Day 5 forecast lead time.*

From the figures, for GPM, it can be seen that the forecast is reliable for all thresholds. At very low probability values the reliability curve lies close to the line of perfect reliability. At probability values above 0.2, the curve shows over-forecasting for all three probability thresholds. Over-forecasting is more in Figure 8(a) for a precipitation threshold of 2.5mm/day than the other two thresholds. The sharpness diagram of Figure 8(a) also shows that on most occasions forecast probability is 100%. Though there is over-forecasting for all three precipitation thresholds it is clear from the figures that higherforecast probability is corresponded by higher observed relative frequency. The over-forecasting in the reliability diagram is not fully due to the limitation of the forecasting system, the small sample size of the data also has contributed significantly to it (Richardson, 2001).

In the case of Merged data as verifying observation, Figure 8(a) shows that for light precipitation the resolution and reliability term both are low compared to the curves obtained using GPM data. Like GPM, the reliability curve for Merged data also showing a large number of forecasts with 100% probability. The nature of the reliability curves for precipitation thresholds of 25 mm/day and 45 mm/day is the same as those obtained in the case of GPM data but the performance of the model using Merge data as verifying observation has declined as can be seen from lower values of resolution and higher values of reliability in Figures 8(b) and 8(c), respectively.

3.5. Relative Operating Characteristics (ROC) Curve

The Relative Operating Characteristic (ROC) plot helps us distinguish between the occurrence and non-occurrence of events. It is a discrimination-based graphical display of forecast verification. The ROC diagram was first introduced into the world of meteorology by Mason (1982).

The ROC curve plots a graph of hit rate (HR) against false alarm rate (FAR). Hit rates and false alarm rates are calculated using the probability thresholds to determine a ‘yes’- ‘no’ forecast when charted in a contingency table (Table 1). The false alarm rate and hit rate are calculated as:

FAR = No. of false alarms/ total number of ‘No’ observations

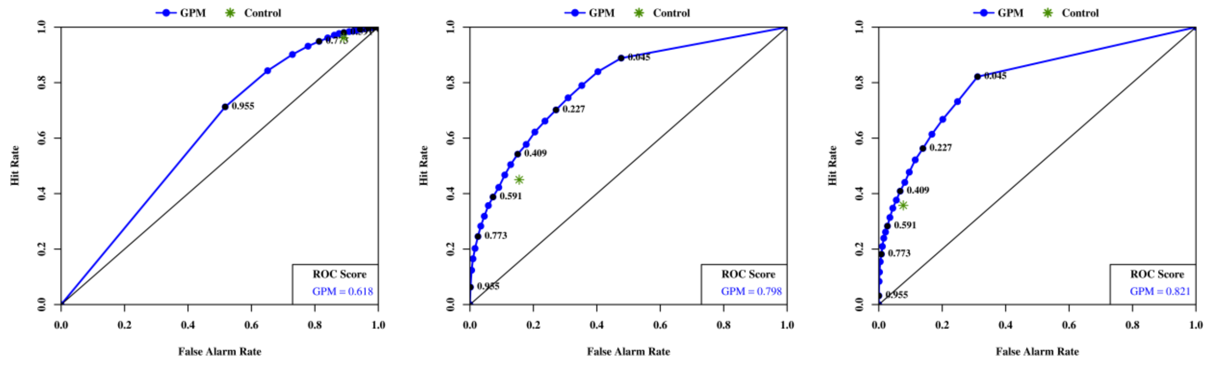
HR = No. of hits / total number of ‘Yes’ observations

		Observation		
		Yes	No	
Forecast	Yes	**Hit**	**False Alarm**	Forecast Yes
	No	**Miss**	**Correct Negatives**	Forecast No
		Observed Yes	Observed No	Total

Table 1. Contingency Table for Forecast and Observation.

Both HR and FAR will be high for low probability threshold values. On increasing the probability threshold value, the probability of correctly predicting the event decreases but simultaneously the false alarm rate also decreases. If the rate of decrease in hit rate is less than that of the false alarm rate, the model prediction can be called as skilful. The diagonal line in an ROC curve shows an equal rate of change of hit rate and false alarm rate, thus considered as no skill line. The area under the ROC curve also serves as another metric of study for the purpose of verification. The area reflects the measure of the discrimination quality of the forecast system. It can have a maximum value of 1 for a perfect forecast and a minimum of 0. However, an area of 0.5 shows no skill.

ROC curves for day-5 forecast of precipitation exceeding 2.5, 25 and 45 mm/day over South Bhutan using GPM data are shown in Figures 9(a), 9(b) and 9(c) respectively. The green asterisks in these figures indicate the HR and FAR of control forecast. For precipitation exceeding 2.5 mm/day the curve in Figure 9(a) shows very high values of HR and FAR. This specifically does not show any remarkable skill of the system. High values of both HR and FAR are due to the fact that the model has a large wet bias for low precipitation hence it predicts a yes forecast more often than actual, which on the other hand increases the false alarm rates as well. The area under the curve is nearly 0.618 which is indicative of some skill. NEPS-G shows much better skill for forecast of precipitation exceeding 25 mm/day (Figure 9(b)). In this case, the hit rates are very high as compared to the false alarm rates. Hence, the curve is inclined towards the vertical axis. The false alarm rate in this case is nearly zero when 90% of the members predict the event. The area under the curve is 0.798 which is indicative of a skillful forecast.



(a) (b) (c)

Figure 9. *The ROC curves for South Bhutan using GPM data for precipitation exceeding (a) 2.5mm/day (b) 25mm/day and (c) 45 mm/day for Day 5 forecast lead time. The green asterisk indicates the control forecast.*

For precipitation exceeding 45 mm/day, the curve shows high hit rates and fairly low rate of false alarms. The curve almost coincides with the y-axis initially showing perfect hits. The truncated shape of ROC at the lower probability threshold end indicates the need of a larger ensemble size for skillful forecast of heavy rainfall events. The area under the curve is 0.821, which is by far the highest area under the curve.

The fact that the green asterisk lies below the ROC curve for the probabilistic forecast in all the figures indicates that the probabilistic forecast has better skill than the control forecast for all three precipitation thresholds.

ROC curves for day-5 forecast of precipitation exceeding 2.5, 25 and 45 mm/day over South Bhutan using Merged data are shown in Figures 10(a), 10(b) and 10(c) respectively. These figures show a fairly good forecast skill of NEPS-G, for higher thresholds in particular, similar to that obtained using GPM data but slightly better as seen by the larger areas under the curves in case of low thresholds. For precipitation threshold of 2.5 mm/day the ROC curve in Figure 10(a) indicates high values of both HR and FAR, quite similar to the case using GPM data (Figure 9(a)). In this case also false alarm rate is more than 50% even when more than 95% of ensemble members predict precipitation. The area under the curve is nearly 0.625 which is slightly higher than the value obtained using GPM data.

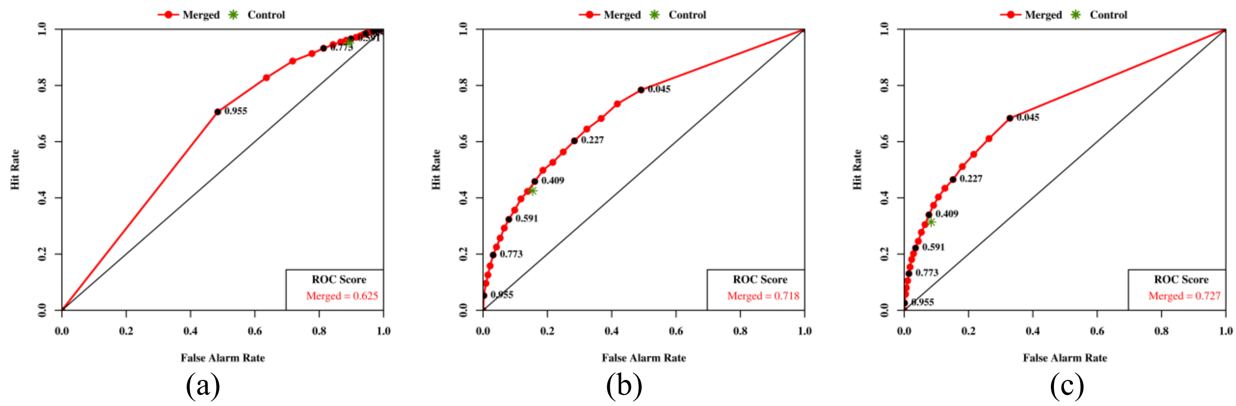


Figure 10. *The ROC curves for South Bhutan using Merged data for precipitation exceeding (a) 2.5mm/day (b) 25mm/day and (c) 45 mm/day for Day 5 forecast lead time. The green asterisk indicates the control forecast.*

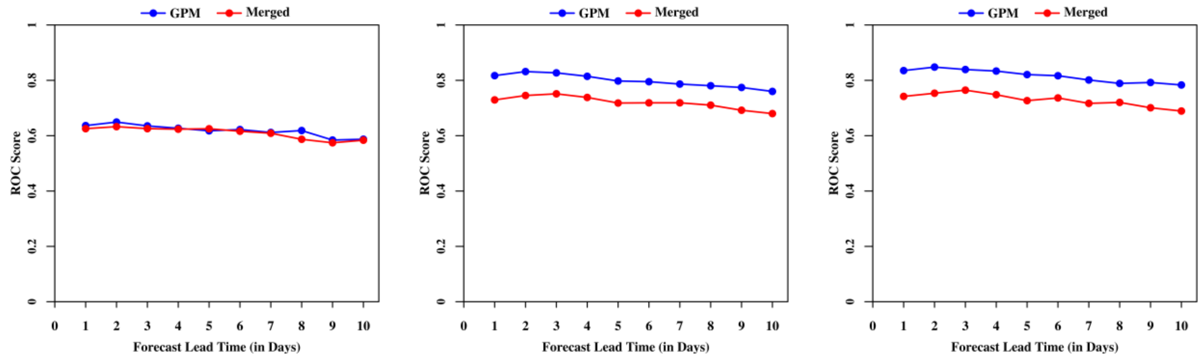
For precipitation threshold greater than 25 mm/day as shown in Figure 10(b), it is a much better skilled forecast. The hit rates are high and false alarm rates are low. Hence, the curve is inclined towards the vertical axis. So the forecast is skillful but the area under the curve (0.718) is not as high as it was obtained (0.798) using GPM data as verifying observation.

The ROC curve for precipitation threshold of 45 mm/day shows the best discrimination property as it was shown in the case of verification using GPM data. The area under the curve is 0.727 which indicates a more skilful forecast than the other two precipitation thresholds but the skill is not as good as was obtained verifying against GPM data. In this case also the green asterisk representing the control forecast lies below the ROC curves for all three precipitation thresholds indicating better discrimination properties of the probabilistic forecast than the deterministic forecast.

Area under ROC Curve using GPM and Merged Data for South Bhutan

Since the forecast skill of a forecasting system reduces with forecast lead time the area under the ROC curve (also called ROC score) also should decrease with forecast lead time. Figure 11 shows that ROC score reduces with forecast lead time for all three precipitation thresholds. This is true for both the cases of verifications using GPM and Merged data.

As seen from Figure 11 ROC score obtained using GPM data is greater than the merged data at all forecast lead times for precipitation exceeding 25 and 45 mm/day whereas, for precipitation greater than 2.5 mm/day, both GPM and Merged datasets are showing identical ROC score at most of the lead times. Both for GPM and Merge data, highest ROC score is obtained for threshold value of 45 mm/day and the lowest ROC score for threshold value of 2.5 mm/day at all forecast lead times. The ROC score for the control forecast is the area under the straight lines connecting the green asterisk (Figures 9 and 10) with the points HR=0, FAR=0 (i.e., the origin) and HR=1, FAR=1 (i.e., the top left corner point). Since the green asterisk always lies below the ROC curve of the probabilistic forecast the ROC score of the control forecast is always smaller than that of the probabilistic forecast of NEPS-G indicating the superiority of the probabilistic forecast of an ensemble prediction system.



(a) (b) (c)

Figure 11. *Variation of ROC score with forecast lead time for South Bhutan using GPM (blue) and Merged (red) data as verifying observation for precipitation exceeding (a) 2.5 mm/day, (b) 25 mm/day and (c) 45 mm/day.*

4. Summary and Conclusion

Bhutan is a landlocked country located in the eastern Himalayan region and is a member of BIMSTEC. BCWC is an organization under BIMSTEC and established at NCMRWF, India. One of the aims of BCWC is to promote collaborative research among the scientists of member countries in the field of weather prediction and climate change. In the present study, researchers of Bhutan and India have carried out the verification of probabilistic precipitation forecast over Bhutan provided by the 23-member global ensemble prediction system of NCMRWF (NEPS-G). The GPM IMERG Final run precipitation data (denoted as GPM) and the satellite-gauge merged data (denoted as Merged data) prepared at NCMRWF, both of 0.1^0 resolution, have been used as the verifying data. The standard probabilistic verification metrics like RMSE-spread relationship, Brier Skill Score, Rank Histogram, Reliability diagram and Relative Operating Characteristic have been used to validate the daily precipitation forecast of Monsoon Months (June, July, August and September) of 2020.

RMSE value for North Bhutan is lower than that of South Bhutan. Since the verifying data over North Bhutan is not reliable (due to sparse gauge density and erroneous GPM data) the low value of RMSE may not be indicative of good forecast over the region. Moreover, contrary to the usual characteristic of forecast the RMSE over North Bhutan does not increase with forecast lead time. However, the probabilistic forecast is under-dispersive over both regions and spread increases with lead time.

Due to the unreliability of the verifying data over North Bhutan, all the other verification metrics have been determined over South Bhutan only. Since the area of verification reduces the results suffer from obvious problem of small sample size. The Brier Skill Score is negative for light precipitation both for GPM and Merged data. This may be due to the well-known positive bias of model forecast for light rainfall category. Model shows positive skill when verified against GPM data for heavier rainfall categories (>25mm/day and >45 mm/day) nearly for all forecast lead days but when verified against Merged data model skill is positive only for moderate precipitation (>25 mm/day) during first four forecast lead days. The better forecast skill achieved using GPM data may be due to the fact that both model and GPM under-estimate the heavier rainfall. For all three forecast categories, model shows usual characteristic of decreasing skill with increasing forecast lead time.

The rank histogram shows that the forecast is under-dispersive both for GPM and Merged data but the distribution is more uniform against Merged data. The day-5 reliability curves for all three categories show that the forecast is reliable. The ensemble prediction system over-forecasts precipitation of all three categories. The small sample size of the data may also have contributed to the over-forecasting nature of the results. The reliability is found to be better when the forecast is verified against GPM data. The ROC curves show that the forecast has good skill in discriminating events from non-events for all three categories but it is most skilful for heavy precipitation category (>45 mm/day). The positive bias in light precipitation forecast is evident from the nature of the curve which shows high false alarm rate at high probability threshold value. The area under ROC curve or ROC score reduces with forecast lead time for all three categories. Highest ROC score is achieved for heavy precipitation category and the score is better when forecast is verified against GPM data.

References

Atger, F. (2004). Estimation of the reliability of ensemble-based probabilistic forecasts. *Quarterly Journal of the Royal Meteorological Society*, *130*(597 PART B), 627–646. https://doi.org/10.1256/qj.03.23

Bishop, C. H., Etherton, B. J., & Majumdar, S. J. (2001). Adaptive Sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical Aspects. *Monthly Weather Review*, *129*, 420–436.

https://doi.org/https://doi.org/10.1175/1520-0493(2001)129<420:ASWTET>2.0.CO;2

Chakraborty, P., Sarkar, A., Bhatla, R., & Singh, R. (2021). Assessing the skill of NCMRWF global ensemble prediction system in predicting Indian summer monsoon during 2018. *Atmospheric Research*, *248*. https://doi.org/10.1016/j.atmosres.2020.105255

Cressman, G. P. (1959). An Operational Objective Analysis System. *Monthly Weather Review*, *87*(10),367–374.https://doi.org/https://doi.org/10.1175/1520-0493(1959)087%3C0367:AOOAS%3E2.0.CO;2

Hamill, T. M. (2001). Interpretation of rank histograms for verifying ensemble forecasts. *Monthly Weather Review*, *129*(3), 550–560. https://doi.org/10.1175/1520-0493(2001)129<0550:IORHFV>2.0.CO;2

Hou, A. Y., Kakar, R. K., Neeck, S., Azarbarzin, A. A., Kummerow, C. D., Kojima, M., Oki, R., Nakamura, K., & Iguchi, T. (2014). The global precipitation measurement mission. *Bulletin of the American Meteorological Society*, *95*(5), 701–722. https://doi.org/10.1175/BAMS-D-13-00164.1

Huffman, G. J., Bolvin, D. T., Braithwaite, D., Hsu, K. L., Joyce, R. J., Kidd, C., Nelkin, E. J., Sorooshian, S., Stocker, E. F., Tan, J., Wolff, D. B., & Xie, P. (2020). Integrated Multi-satellite Retrievals for the Global Precipitation Measurement (GPM) Mission (IMERG). In *Advances in Global Change Research* (Vol. 67, pp. 343–353). Springer. https://doi.org/10.1007/978-3-030-24568-9_19

Krishnamurti, T. N., S. Cocke, R. Pasch, and S. Low-Nam, 1983: Precipitation estimates from rain gauge and satellite observations, summer MONEX. FSU Rep. 83-7, Department of Meteorology, The Florida State University, 373 pp

Mamgain, A., Sarkar, A., & Rajagopal, E. N. (2020a). Medium-range global ensemble prediction system at 12 km horizontal resolution and its preliminary validation. *Meteorological Applications*, *27*(1). https://doi.org/10.1002/met.1867

Mamgain, A., Sarkar, A., & Rajagopal, E. N. (2020b). Verification of high resolution (12 km) Global Ensemble Prediction System. *Atmospheric Research*, *236*, 104832. https://doi.org/10.1016/j.atmosres.2019.104832

Mason, I. (1982). A Model for Assessment of Weather Forecasts. *Australian Meteorological Magazine*, *30*, 291–303.

Mitra, A. K., Bohra, A. K., Rajeevan, M. N., & Krishnamurti, T. N. (2009). Daily Indian precipitation analysis formed from a merge of rain-gauge data with the TRMM TMPA satellite-derived rainfall estimates. *Journal of the Meteorological Society of Japan*, *87 A*, 265–279. https://doi.org/10.2151/jmsj.87A.265

Mitra, A. K., Das Gupta, M., Singh, S. V, & Krishnamurti, T. N. (2003). *Daily Rainfall for the Indian Monsoon Region from Merged Satellite and Rain Gauge Values: Large-Scale Analysis from Real-Time Data*.

Palmer, T. (2019). The ECMWF ensemble prediction system: Looking back (more than) 25 years and projecting forward 25 years. *Quarterly Journal of the Royal Meteorological Society*, *145*(S1), 12–24. https://doi.org/10.1002/qj.3383

Reddy, M. V., Mitra, A. K., Momin, I. M., & Krishna, U. V. M. (2022). How Accurately Satellite Precipitation Products Capture the Tropical Cyclone Rainfall? *Journal of the Indian Society of Remote Sensing*. https://doi.org/10.1007/s12524-022-01572-1

Reddy, M. V., Mitra, A. K., Momin, I. M., Mitra, A. K., & Pai, D. S. (2019). Evaluation and inter-comparison of high-resolution multi-satellite rainfall products over India for the southwest monsoon period. *International Journal of Remote Sensing*, *40*(12), 4577–4603. https://doi.org/10.1080/01431161.2019.1569786

Richardson, D. S. (2001). Measures of skill and value of ensemble prediction systems, their interrelationship and the effect of ensemble size. *Quarterly Journal of the Royal Meteorological Society*, *127*(577), 2473–2489.

Sanchez, C., Williams, K. D., & Collins, M. (2016). Improved stochastic physics schemes for global weather and climate models. *Quarterly Journal of the Royal Meteorological Society*, *142*(694), 147–159. https://doi.org/10.1002/qj.2640

Shanker, G., Sarkar, A., Mamgain, A., Prasad, S. K., Bhatla, R., & Mitra, A. K. (2022). Contribution of lagged members to the performance of a global ensemble prediction system. *Atmospheric Research*, *280*. https://doi.org/10.1016/j.atmosres.2022.106451

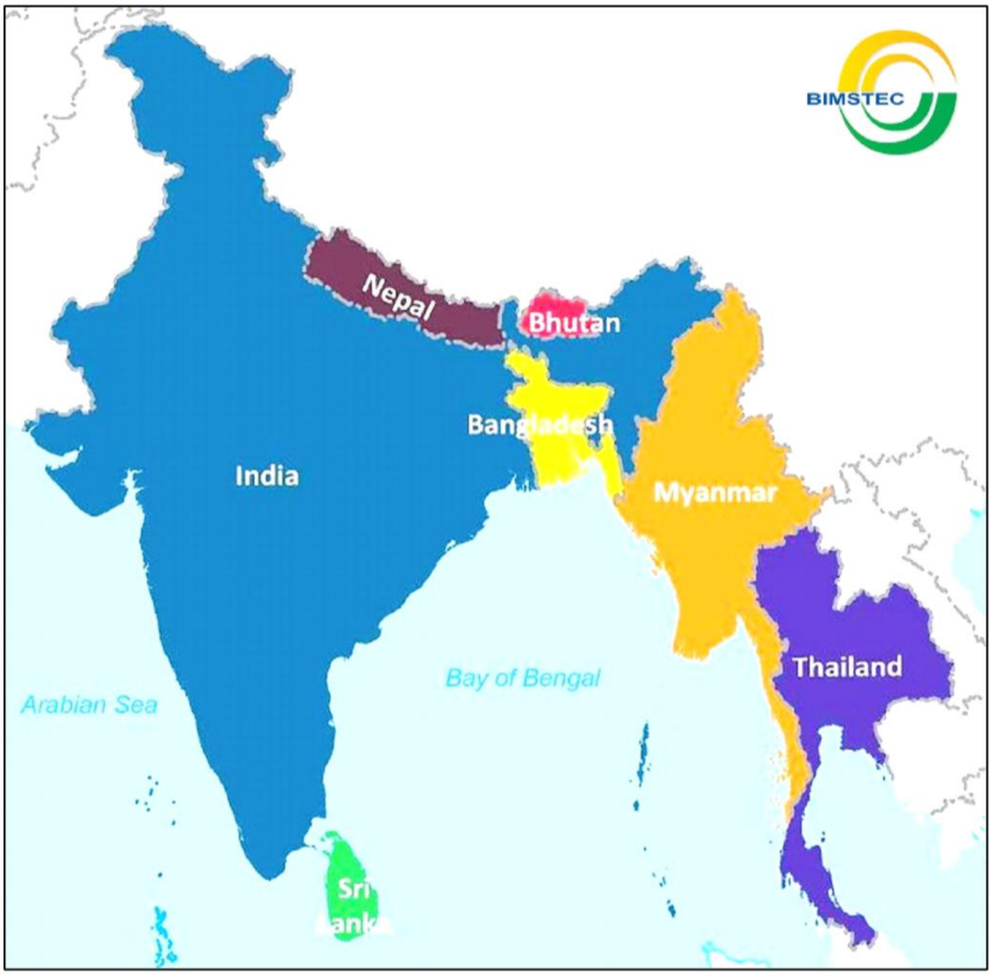
Tennant, W. J., Shutts, G. J., Arribas, A., & Thompson, S. A. (2011). Using a stochastic kinetic energy backscatter scheme to improve MOGREPS probabilistic forecast skill. *Monthly Weather Review*, *139*(4), 1190–1206. https://doi.org/10.1175/2010MWR3430.1

Tripoli, G. J., & Krishnamurti, T. N. (1975). Low-Level Flows over the GATE Area during Summer 1972. *Monthly Weather Review Review*, *103*(3), 197–216. https://doi.org/https://doi.org/10.1175/1520-0493(1975)103%3C0197:LLFOTG%3E2.0.CO;2

Wilks, D. S. (2019). Statistical Methods in the Atmospheric Sciences. In *Statistical Methods in the Atmospheric Sciences*. Academic Press, London.

BCWC/RR/01/2024

BCWC RESEARCH REPORT



Contact
Head, NCMRWF
Phone: +91-120-2419401
Fax: +91-120-2419484 [View Route Map]

Bangladesh Bhutan India Myanmar Nepal Sri Lanka Thailand

BIMSTEC Centre for Weather and Climate (BCWC)

National Centre for Medium Range Weather Forecasting,
Ministry of Earth Sciences, Government of India
A-50, Sector-62, NOIDA-201 309, INDIA