Verification and Bias Correction of Rainfall and Temperature Forecasts over the Babai River Basin of Nepal

Dilip Kumar Gautam[1]*, and Raju Dhar Pradhananga[2]

[1]*Regional Integrated Multi-Hazard Early Warning System (RIMES), Bangkok, Thailand*
[2]*Department of Hydrology and Meteorology, Government of Nepal .*

(Received 01 November 2023; Accepted 21 September 2024)

ABSTRACT

Weather Research & Forecasting model is known to exhibit systematic biases for weather variables. These biases need to be post-processed to get an optimal result before applying the weather forecast data in hydrological modeling or similar applications. In this paper, we examined the performance of Weather Research & Forecasting Model forecasts for rainfall and temperature over the Babai River basin of Nepal considering various performance indicators using statistical approach. The model was able to capture the rainfall event forecast (Rain/No Rain) sufficiently. However, the model showed poor skill in forecasting the amount and over-forecasted the rainfall. The multi-category (No Rain, Light Rain, Moderate Rain and Heavy Rain) verifications showed over-forecast in light rain and moderate rain categories and under-forecast in no rain and heavy rain categories. We examined various bias correction schemes such as distribution-derived transformations, parametric transformations and nonparametric transformations to get de-biased results. The empirical quantile mapping is the best scheme for bias correction of both rainfall and temperature. In case of temperature, the linear transformation, robust quantile mapping and smoothing spline schemes also performed well.

Keywords: Precipitation, Temperature, WRF, Forecast Verification and Bias Correction

1. Introduction

The summer monsoon is the main rainfall season in Nepal. About 79 % of the total annual rainfall occurs during this season (DHM, 2017). According to the Department of Hydrology and Meteorology of Nepal, the normal date of onset of monsoon in Nepal is 13th of June and its normal date of withdrawal is 2nd of October. Although the monsoon normally stretches from June to September, its actual onset as well as the withdrawal dates are uncertain. Both the dates of onset and withdrawal of summer monsoon are found to be delayed in recent years (Gautam and Regmi, 2013). After the onset of monsoon, it will cover the whole country within a week. The onset of monsoon is also known as burst of monsoon because there is sudden change from hot and dry weather to wet and humid weather. The onset showers can last for days after which it takes a steady pattern of rain for few hours most days. The pattern of monsoon rainfall over the Indian subcontinent is greatly influenced by the position of monsoon trough, frequency and tracks of monsoon depressions originating in the Bay of Bengal and the complex topography it follows afterward.

DOI: https://doi.org/10.3126/jhm.v12i1.72624

**Corresponding author:* Dilip K. Gautam, dilipgautam65@gmail.com

The intensity and duration of monsoon are not uniform from year to year. In some years it rains too much causing floods and landslides while in other years it rains fewer causing droughts. In some years, average monsoon rainfall is good but timing may not be proper and, in some years, its daily distribution may be largely skewed. This is called variability of monsoon. On the one hand, monsoon supplies us with fresh water for agricultural production, hydropower generation, and groundwater recharge, on the other hand, it causes floods resulting in huge financial losses, damage to lives and properties and destruction of the environment. Therefore, it is important to predict and understand monsoon rainfall patterns.

Weather systems in Nepal have so far been little researched by the atmospheric science community and weather observation is largely lacking in coverage and modern technology and the forecast has severely handicapped due to poor performance of computer models in this region. All modern forecasting methods involve observation of current conditions, along with the combination of historical data, scientific methods and computer modeling (Gibilisco, 2005). Numerical weather prediction (NWP) is the forecasting of the weather based on the solutions of mathematical equations by high-speed computers. In this regard, the Advanced Research Weather Research & Forecasting (ARW) model is the state-of-the-art NWP model for the atmospheric circulation system (Dudhia, 2004). It is designed to serve both operational forecasting as well as atmospheric research needs. One of the greatest challenges of NWP is to improve quantitative precipitation forecasting (QPF) significantly (Yates et al., 2006). The weather forecasting models have systematic errors or bias problems that are inherent in model physics, parameterization schemes, initial conditions, resolution of the model, etc. These biases can be reduced using bias-correction schemes (Gudmundsson et al., 2012). The resulting forecast with less forecast error can be fed into the flood forecasting models for the reliable prediction of floods. A reliable forecasting and warning system can save lives and reduce the loss of properties. Bannister et al. (2019) investigated the ability of bias-corrected WRF model output at 5 km grid spacing to reproduce the spatiotemporal variability of precipitation for the Beas and Sutlej River basins in the Himalaya, measured by 44 stations spread over the period 1980 to 2012 and found that the raw (uncorrected) model output generally underestimated annual, monthly, and daily precipitation amounts. However, applying a nonlinear bias-correction method to the model output resulted in much better results, which were superior to precipitation estimates from reanalysis and two gridded datasets. These findings highlight the difficulty in using current gridded datasets as input for hydrological modeling in Himalayan catchments, suggesting that bias-corrected high-resolution regional climate model output is in fact necessary. In this study, we verified the rainfall and temperature forecasts by WRF model over the Babai River Basin (BRB) in western Nepal. We also tested various bias-corrections schemes and identified the best schemes for correcting biases in the rainfall and temperature forecasts. No such study has been conducted before for verification and bias correction of WRF model forecasts over the river basins of Nepal. Babai river is a rain-fed river. BRB is one of the most flood-prone river basins in Nepal. During the 2014 floods in Western Nepal, despite an established early warning system, 31 people lost their lives along the Babai River in Bardiya district (Shrestha et al., 2021). Flood forecasting over the BRB could be significantly improved if bias-corrected rainfall and temperature forecasts are employed in the flood forecasting model.

2. Study Area

BRB is a medium-sized river basin located in the western part of Nepal. It covers three districts namely Salyan, Dang and Bardiya. Its catchment area is about 3513 km^2. It originates in the Siwalik Mountain range, flows northwestward parallel with the Bheri River and then southward into Terai plain passing through the Bardiya National Park. The basin has a very steep gradient in its upper course which becomes gentle as the river enters the Dun Valley. Here it is joined by small tributaries flowing in both from the lower Himalayas in the north and the Siwalik hills in the south. The elevation ranges from around 100 m at the Nepal-India border to around 2500 m in the northern part of the mountains. The locations of the Babai basin, precipitation stations and temperature stations are shown in Fig. 1.

3. Methodology

The methodology consists of data pre-processing, forecast verification and bias correction.

3.1. Data Pre-processing

The observed daily data from 16 precipitation stations and 4 temperature stations (Ghorahi, Tulsipur, Salyan Bazaar and Rani Jaruwa) as shown in Fig. 1 during the study period of 2008 - 2013 was obtained from the Department of Hydrology and Meteorology, Nepal. The daily rainfall data are recorded at 03:00 UTC. Some of the stations are upgraded to Automatic Weather Station (AWS). The missing daily rainfall and temperature data were filled by AWS data of the same station where available. The remaining missing data were filled using data of the neighboring stations by a normal-ratio method ((Paulhus and Kohler, 1952); (Young, 1992)). This method is preferred in the mountainous regions where the annual average rainfall differs considerably between locations. The consistencies of data filling were checked by using double mass curves.

WRF model forecast is provided for 9 km grid resolution, i.e., the forecast is averaged over 81 sq. km. area, which is a large area, almost equal to a sub-basin. The rainfall may vary significantly over this area. Hence, the rainfall verification and bias correction were done on the sub-basin scale.

Since the temperature doesn't vary significantly over the area covered by a single grid, the temperature verification and bias correction were done on the station scale. The average observed rainfall for each sub-basin was calculated using the Thiessen polygon method. The Thiessen polygon method is more suitable in moderately rugged areas. This method is useful for areas, which are more or less plain and are of intermediate size (500 to 5000 km^2) and when there are a few rain gauge stations compared to the size of the basin (http://ecoursesonline.iasri.res.in/mod/page/view.php?id=2212). With the help of ArcGIS software, the basin was divided into six sub-basins W60, W80, W90, W140, W180 and W190, representing watersheds from upstream to downstream (see Fig. 1).

We employed the forecast data of the WRF model run by the Regional Integrated Multi-Hazard Early Warning System for Africa and Asia (RIMES) in Bangkok, Thailand (https://www.rimes.int/pillar2). The WRF model is a state-of-the-art mesoscale numerical weather prediction

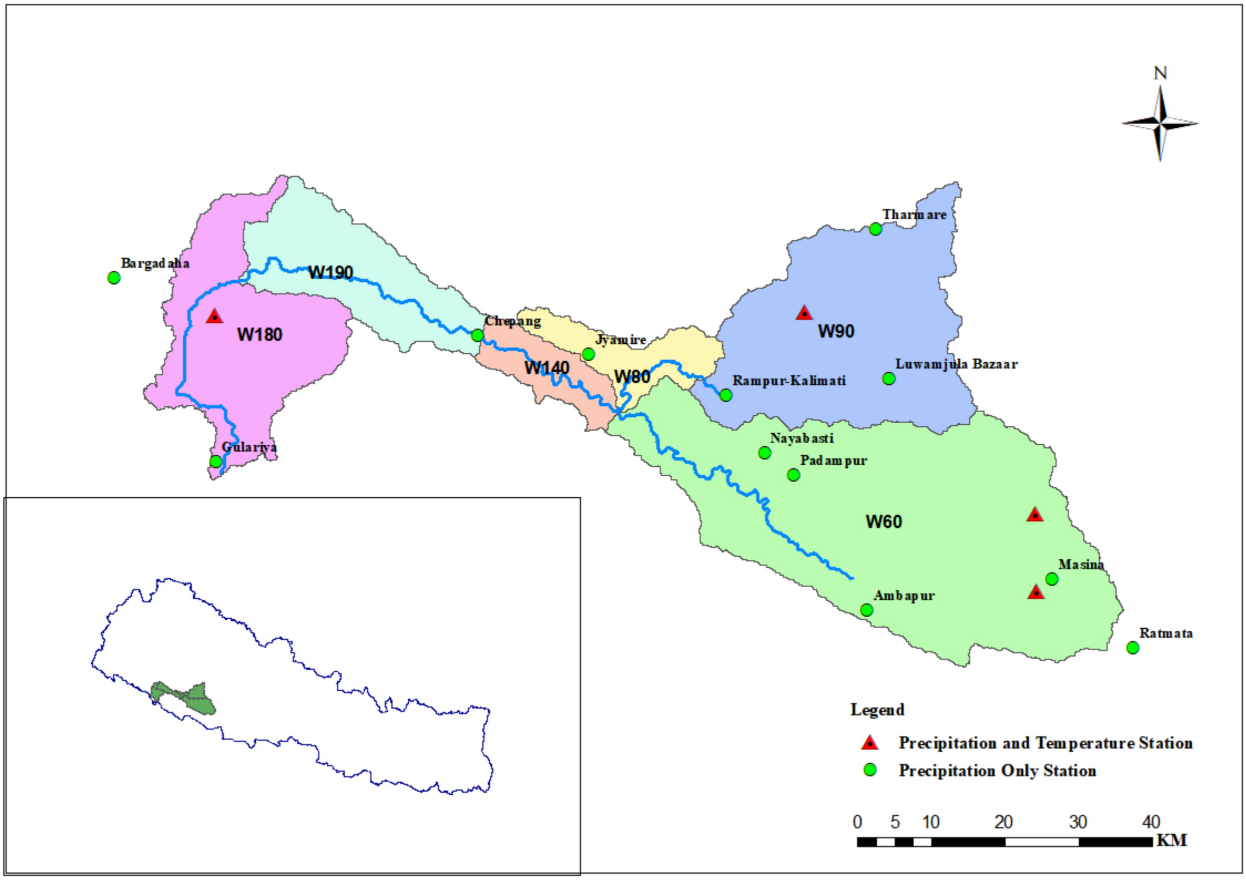


Figure 1. Location map of Babai River Basin in western Nepal along with precipitation and temperature stations'.

system designed for both atmospheric research and operational forecasting applications (https://www.mmm.ucar.edu/models/wrf). The current operational model setup for the WRF model at RIMES covers the region from 20°E to 150°E longitude and 16°S to 50°N latitude. The model is configured to run with National Center for Environmental Prediction (NCEP) Global Forecasting System (GFS) data downloaded for 12 UTC initial conditions. The details of the model parameters are presented in Table 1. The 24-category USGS land-use and topography data sets were used for interpolating topography and land use with a spatial resolution of 2' for domain. The NCEP Global Forecast System (GFS) data of resolution 1°× 1°with Grib2 format was used as input for initial and boundary conditions to the model which was taken at six hourly intervals. The output has been generated from the model every 180 minutes.

3.2. Forecast Verification

Verification is the process of comparing forecasts to relevant observations, which measures the quality of forecasts (Fowler et al., 2012). The verification of forecasts is important(a) to improve model forecast, (b) to improve decision making, (c) to understand model biases, and (d) to make choice of a better model or better model configuration.

These methods include continuous verification, dichotomous (binary) verification, multi-category verification, visual inspection of maps and plots (such as time series, scatterplots, quantile plots, density plots, and box plots), and spatial plots of forecast errors. Further details on these methods are available at http://www.cawcr.gov.au/projects/verification/#Methods_for_spatial_forecasts.

Table 2 below presents the performance indicators of continuous forecast verification.

Performance indicators of dichotomous (binary) and multi-category forecast verification are calculated using a contingency table as given in Table 3.

3.3. Bias Correction

Nepal has complex topography with plains in the southern belt, high mountains in the northern belt and middle hills and valleys in the middle belt. The bias in precipitation is found to vary spatially. Therefore, the bias corrections were carried out for each of 6 sub-basins of Babai River basin separately. For our study 8 bias correction schemes

Table 1. RIMES Operational Model Parameter Set

SN	Parameters	Values
1	Model domain	20°E to 150°E and 16°S to 50°N
2	Grid Resolution	9 km x 9 km
3	Projection	Mercator
4	Topographical data	USGS (2m)
5	No of grid points X direction	1470
6	No. of grid points in Y-direction	870
7	Forecast Interval	6 hourly
8	Time Step	45 s
9	No of vertical levels	27
10	Micro Physics option	5 (Ferrier (new Eta))
11	Cumulus Scheme	1 (Kain-Fritsch)
12	Forecast Lead time	84 Hours

were chosen. The first four were parametric transformation schemes (qm1: scale, qm2: linear, qm3: power and qm4: exponential asymptotic) the fifth one was a distribution-derived transformation scheme (qm5: Bernoulli-gamma) and the last three were non-parametric transformation schemes (qm6: Robust Empirical Quantiles, qm7: Empirical Quantiles, and qm8: Smooth Spline).

Parametric Transformation: The quantile-quantile relation of observed and modelled value is fitted by using the transformation to adjust the distribution of the modelled data to match the distribution of the observations. The following parametric transformations were used for the study.

Scale:

$$\hat{P}_o = b \times P_m$$

Linear:

$$\hat{P}_o = a + b \times P_m$$

Power:

$$\hat{P}_o = b \times P_m^c$$

Exponential Asymptotic:

$$\hat{P}_o = (a + b \times P_m) \times \left(1 - e^{-\frac{P_m}{\tau}}\right)$$

Where a, b, c, and τ are constants, P_m is the model precipitation, and o is the best estimate of the observed precipitation.

Distribution Derived Transformation: The Bernoulli-Gamma Transformation was also tested for bias correction. This transformation is a mixture of Bernoulli and Gamma distributions. The parameters of these distributions are estimated by maximum likelihood methods for both P_o and P_m independently (Yates et al., 2006).

Non-Parametric Transformation: The following non-parametric transformations were used for the study.

Robust Empirical Quantiles: It estimates the values of the quantile-quantile relation of observed and modelled time series for regularly spaced quantile using local linear least square regression and performs quantile mapping by interpolating the empirical quantiles.

Empirical Quantiles: It estimates values of the empirical cumulative distribution function of observed and modelled times series for regularly spaced quantiles and uses these estimates to perform quantile mapping.

Smoothing Spline: It uses the spline function to adjust the distribution of the modelled data to match the distribution of the observation.

Verification (Gilleland, 2010) and qmap (Gudmundsson, 2016) packages in R (R Core Team, 2013) have been utilized for forecast verification and bias correction in this study.

4. Results

4.1. Verification of Rainfall Forecast

4.1.1) Continuous Forecast Verification

This method was used to evaluate how the values of forecasts differ from the values of observations. The amounts of rainfall forecasted by the model for all sub-basins were compared with the observed rainfalls. PBIAS and skill scores were computed to evaluate the performance of the model. The model showed very poor skill in forecasting rainfall amounts for all subbasins. As shown in Fig. 2(a) all sub-basins have negative skill scores indicating poor forecasts for all 3 days forecasts. The PBIAS in Fig. 2(b) indicates that the model over-forecasted rainfall in sub-basins W90, W180, and W190 which lie in Terai region whereas it under-forecasted rainfall in W60, W80 and W140 sub-basins which lie between Siwalik and Mahabharat range.

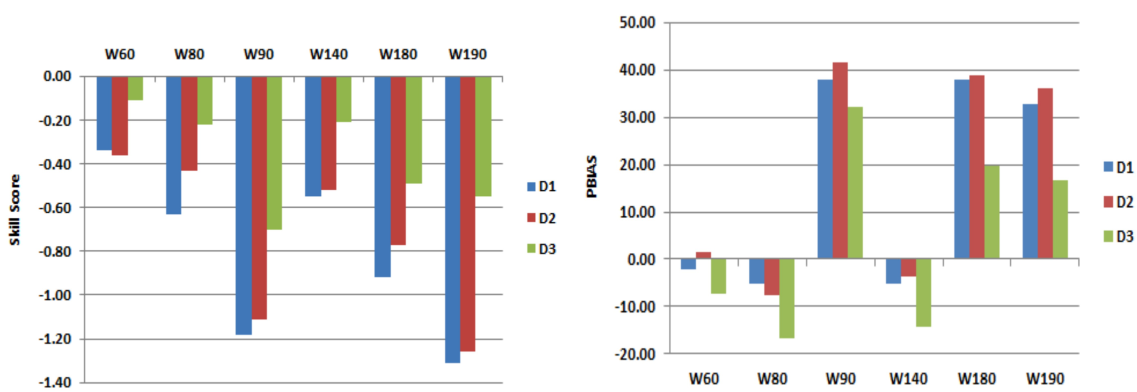
4.1.2) Binary Forecast Verification

This method was used to evaluate whether an event will happen or not e.g. Rain or No Rain. To verify this type of forecast, we used a contingency table as given in Table 3 that shows the frequency of "yes" and "no" forecasts and occurrences. PC, POD, FAR and BIAS were computed for all sub-basins for 3-day forecasts.

The PC was in the range of 66% - 84% indicating good accuracy of the model in forecasting rainfall events. The POD (HR) was in the range of 87% - 95% indicating good detection of rainfall events by the model. The FAR was in the range of 9% - 38%. The BIAS Score was greater than one in all basins indicating an over-forecast of rainfall events. Besides, the performance of model forecasts slightly degraded with the increase in length of forecasts,

Table 2. Performance indicators of continuous forecast verification

SN	Method	Formula	Description
1	Mean Error (ME)	$\text{ME} = \frac{\sum(F_i - O_i)}{n}$	Where F is forecast and O is observation. The perfect score is 0, and the range varies from $-\infty$ to $+\infty$. ME $>$ 0 indicates over-forecasting, and ME $<$ 0 indicates under-forecasting. It measures bias.
2	Mean Absolute Error (MAE)	$\text{MAE} = \frac{\sum\lvert F_i - O_i\rvert}{n}$	The perfect score is 0, and the range varies from 0 to $+\infty$. It gives the average magnitude of errors in a given set of forecasts and measures accuracy.
3	Mean Square Error (MSE)	$\text{MSE} = \frac{\sum(F_i - O_i)^2}{n}$	The perfect score is 0, and the range varies from 0 to $+\infty$. It gives error variance and measures accuracy.
4	Percent Bias (PBIAS)	$\text{PBIAS} = \frac{1}{n}\sum\left(\frac{F_i - O_i}{O_i} \times 100\right)$	The perfect score is 0. A positive value indicates a tendency of model overestimation, and a negative value indicates a tendency of model underestimation.
5	Skill Score (SS)	$\text{SS} = 1 - \frac{\text{MSE}_{\text{forecast}}}{\text{MSE}_{\text{reference}}}$	It measures the relative improvement of the forecast over some reference forecast. The perfect score is 1, and the range varies from $-\infty$ to 1. Zero indicates no improvement over the reference forecast.



(a) (b)

Figure 2. (a) Skill Score for sub-basins and (b) Percent Bias for sub-basins.

Table 3. Confusion Matrix with Event Forecast and Marginal Totals

		Event Observed		
		YES	**NO**	**Marginal Total**
Event Forecast	**YES**	a	b	$a+b$
	NO	c	d	$c+d$
	Marginal Total	$a+c$	$b+d$	$a+b+c+d$

where: a = Hits, b = False Alarms, c = Misses, d = Correct Negatives

i.e., Day 1 forecasts were slightly better than Day 2 forecasts and Day 2 forecasts were slightly better than Day 3 forecasts (see Fig. 3(a) - 3(d)).

4.1.3) Multi-Category Forecast Verification

A contingency table, as given in Table 3, showing the frequency of forecasts and observations in the various bins, e.g., no rainfall, light rainfall, moderate rainfall, and heavy rainfall, was prepared to verify multi-category forecasts. The four categories used for rainfall were as follows: No Rain = 0 mm; Light Rain = 0.1 – 10 mm; Moderate Rain = 10.1 – 30 mm; and Heavy Rain = above 30 mm.

In case of heavy rainfall and no rainfall categories, in most of the sub-basins, in general, the model showed under-forecasts. In contrast, in the case of light and moderate rainfall categories, in most of the sub-basins, in general, the model showed over-forecasts. (See Figs. 4(a) – 4(d)).

Table 4. Performance indicators of dichotomous (binary) and multi-category forecast verification

SN	Method	Formula	Description
1	Bias Score or Frequency Bias (BIAS)	BIAS = $\frac{a+b}{a+c}$	BIAS > 1 indicates over-forecasting, and BIAS < 1 indicates under-forecasting. The perfect score is 1, and the range varies from 0 to ∞.
2	Percent Correct (PC)	PC = $\frac{a+d}{n}$	Indicates the overall fraction of correct forecasts. The perfect score is 1, and the range varies from 0 to 1. Strongly influenced by the common category.
3	Probability of Detection (POD)	POD = $\frac{a}{a+c}$	Represents the fraction of predicted "yes" events that occurred. It is sensitive to misses. The perfect score is 1, and the range varies from 0 to 1.
4	False Alarm Ratio (FAR))	FAR = $\frac{b}{a+b}$	Gives the fraction of predicted "yes" events that did not occur. It is sensitive to false alarms, not misses. The perfect score is 0, and the range varies from 0 to 1.

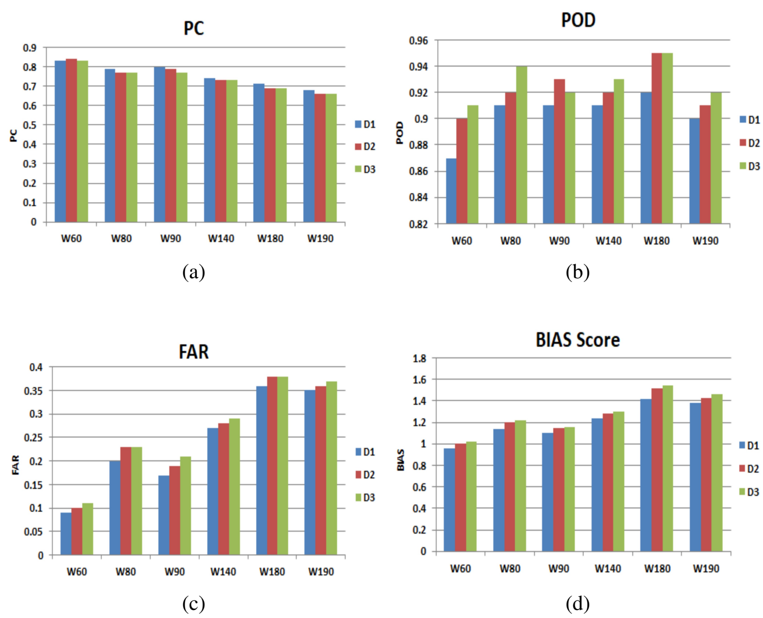


Figure 3. (a) Percent Correct for sub-basins, (b) Probability of Detection for sub-basins, (c) False Alarm Ratio for sub-basins and (d) Bias Score for sub-basins

4.2. Verification of Mean Temperature Forecast

The 3-day temperature forecasts for stations 417 (Rani Jaruwa), 508 (Tulsipur), 511 (Salyan Bazaar) and 515 (Ghorahi) were compared with the observed mean daily temperature. The PBIAS was in the range of -3.1 to +3.23, which is very good result. The PBIAS was positive in stations 417 (Rani Jaruwa) and 515 (Ghorahi) indicating slightly over-forecast whereas it was negative in stations 508 (Tulsipur) and 511 (Salyan Bazaar) indicating slightly under-forecast. The Skill Score was in the range of 0.52 – 0.88 indicating good skill of model in forecasting mean temperatures for the stations under study (See Figs. 5(a) – 5(b)).

To assess the model performance of mean temperature forecasts in multi-categories, verification measures such as bias scores (BIAS) were analyzed. The three categories used for mean temperature analysis were as follows: Category 1 (cold): less than 15 Degree Celsius; Category 2 (warm): 15 – 25 Degree Celsius; Category 3 (hot): above 25 Degree Celsius.

Figs. 6(a) – 6(c) show the Bias Scores for cold, warm and hot conditions. The category 1 (cold condition) is under-forecasted in station 417 (Rani Jaruwa) and over-forecasted in stations 508 (Tulsipur) and 511 (Salyan Bazaar). The category 2 (warm condition) is under-forecasted in station 511 (Salyan Bazaar) and over-forecasted in stations 508 (Tulsipur) and 515 (Ghorahi). The category 3 (hot con-

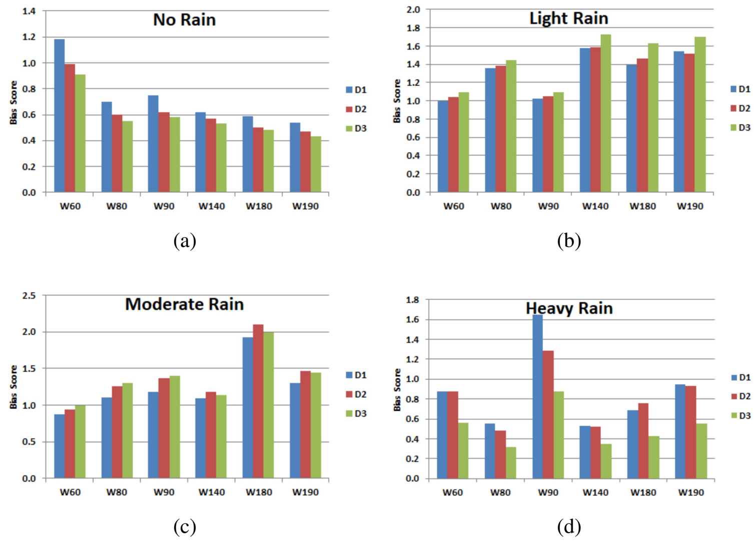


Figure 4. Bias Score for (a) "No Rain", (b) "Light Rain", (c) "Moderate Rain" and (d) "Heavy Rain" categories.

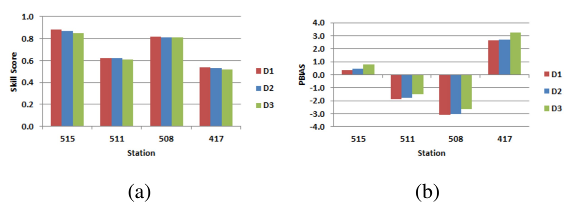


Figure 5. (a) Skill Score for temperature stations and (b) Percent Bias for temperature stations.

dition) is over-forecasted in station 511 (Salyan Bazaar) and under-forecasted in stations 508 (Tulsipur) and 515 (Ghorahi).

4.3. Bias Correction of Rainfall Forecasts

The performances of eight bias correction schemes were evaluated for all 3-days rainfall forecasts for all sub-basins to identify the best bias correction scheme. Figs. 7(a) - 7(c) show the percent bias for Day 1 (D1), Day 2 (D2) and Day 3 (D3) forecasts using different schemes for all sub-basins. The percent bias of model forecast was reduced significantly by all bias correction schemes but empirical quantile scheme has reduced the bias to a minimum. However, other schemes like robust empirical quantile, linear and exponential asymptotic schemes had also reduced the bias significantly.

Upon comparison of the 100% quantile of the bias corrected forecasts from different schemes with 100% quantile of observed rainfall, the bias corrected forecast from the empirical quantile schemes (qm7) was found to be either matching or close to the 100% quantile of the observed

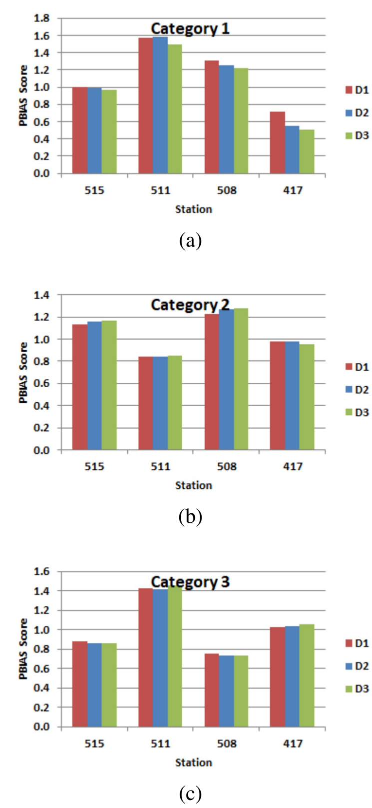
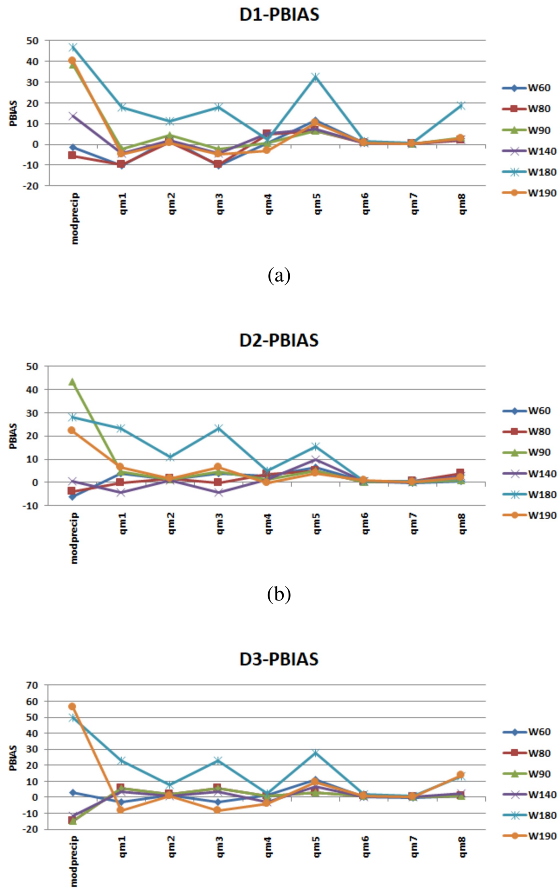


Figure 6. Bias Score for (a) "Cold", (b) "Warm" and (c) "Hot" categories.

rainfall. It suggested that the extreme rainfall cases are well captured by empirical quantile method. Besides, 100%

quantile from other non-parametric schemes such as robust empirical quantile (qm6), smoothing spline (qm8) and parametric scheme such as exponential asymptotic (qm4) were also close to the observed quantile (See Table 5).



(a)

(b)

(c)

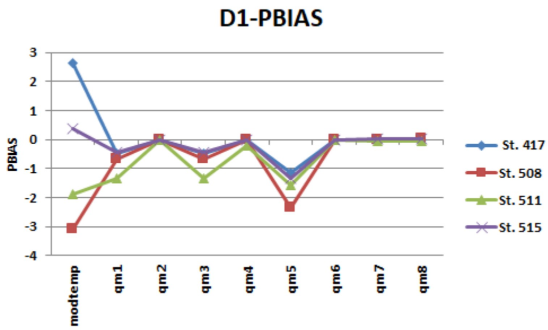
Figure 7. Percent Bias for (a) "Day 1", (b) "Day 2" and (c) "Day 3" rainfall forecast.

4.4. Bias Correction of Temperature Forecasts

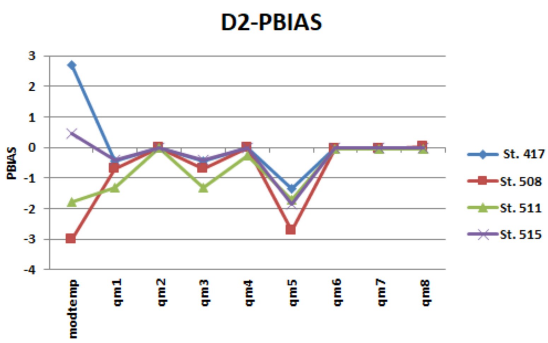
The performances of eight bias correction schemes were evaluated for all 3-days temperature forecasts for all four stations to identify the best bias correction scheme. Figs. 8(a) – 8(c) show the percent bias for Day 1 (D1), Day 2 (D2) and Day 3 (D3) forecasts using different schemes for all stations. The percent bias of mean temperature forecast was less than 5 percent for all stations for 3-days forecast. This was further reduced by all bias correction schemes. However, empirical quantile, robust empirical quantile, smoothing spline, linear and exponential asymptotic schemes have reduced the bias significantly.

Table 5. Comparison of 100% Quantile by Different Schemes with Observed Quantile

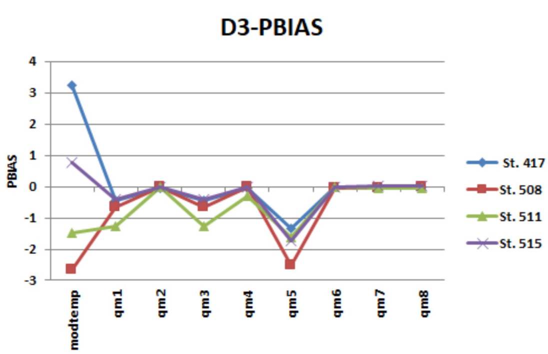
Sub-basin	**W60**			**W80**			**W90**			**W140**			**W180**			**W190**		
Lead Time	D1	D2	D3	D1	D2	D3	D1	D2	D3	D1	D2	D3	D1	D2	D3	D1	D2	D3
Observed	74.5	74.5	74.5	98.2	98.2	98.2	68.5	68.5	68.5	85.2	85.2	85.2	137.0	137.0	137.0	95.6	95.6	95.6
Modelled	98.2	89.2	73.7	130.8	102.2	82.0	97.2	84.8	78.3	128.8	110.2	86.0	157.0	165.2	134.8	144.8	150.5	105.7
qm1	85.0	81.6	78.2	122.5	105.5	99.7	67.4	61.9	60.8	117.8	106.9	100.6	111.0	129.8	116.2	93.6	105.3	93.2
qm2	83.3	81.3	80.9	118.6	106.2	104.9	68.1	66.1	67.1	112.1	104.9	102.7	120.1	141.8	129.7	95.3	107.9	101.6
qm3	85.0	81.6	78.2	122.5	105.5	99.7	67.4	61.9	60.8	117.8	106.9	100.6	111.0	129.8	116.2	93.6	105.3	93.2
qm4	73.3	75.5	75.1	106.0	97.0	104.6	63.8	66.5	68.5	93.6	97.5	103.6	132.7	149.3	141.4	96.1	107.7	102.0
qm5	147.1	141.5	122.0	205.3	166.3	151.8	99.8	90.7	86.9	232.2	200.6	177.9	335.5	405.6	430.3	166.1	196.8	162.2
qm6	71.0	72.6	74.0	96.1	91.8	94.6	62.7	65.7	66.9	89.0	86.0	88.0	118.6	133.7	134.2	86.8	93.4	96.8
qm7	74.5	74.5	74.5	98.2	97.3	98.2	68.5	68.5	68.3	85.2	85.2	85.2	137.0	137.0	137.0	95.6	95.6	95.6
qm8	75.3	75.6	74.5	98.2	97.3	98.2	69.4	68.5	69.2	89.4	86.8	90.4	149.8	221.4	144.8	107.7	199.1	118.5



(a)



(b)



(c)

Figure 8. Percent Bias for (a) "Day 1", (b) "Day 2" and (c) "Day 3" temperature forecast.

5. Conclusion

We employed 3-day forecast data of WRF model run by RIMES and observed data of 16 rainfall stations and 4 temperature stations (Ghorahi, Tulsipur, Salyan Bazaar and Rani Jaruwa) by DHM for the period of 01/01/2008-31/12/2013 for verification and bias-correction of rainfall and temperature over the BRB. The average observed rainfall for each sub-basin was calculated using the Thiessen polygon method.

WRF model forecast is a deterministic forecast. Hence, we employed continuous, binary and multi-category verification methods using different indicators to evaluate the performance of WRF model to forecast rainfall and temperature. We also evaluated the performance of 8 bias correction schemes to identify the best scheme for bias correction of rainfall and temperature forecasts. The rainfall verification and bias correction were done on sub-basin scale and temperature verification and bias correction were done on station scale.

The amounts of rainfall forecasted by the model for all sub-basins were compared with the observed rainfalls. The model showed very poor skill in forecasting the amount of rainfall for all subbasins as indicated by negative skill scores for all 3-day forecasts. The model over-forecasted rainfall over sub-basins of Terai region whereas it under-forecasted over sub-basins of Siwalik and Mahabharat ranges. In general, the model over-forecasted amount of rainfall in the BRB.

The binary verification of forecasting rainfall events (Rain or No Rain) showed good accuracy with the PC ranging from 66% to 84%, POD 87% to 95%, and FAR 9% to 38%. However, the BIAS Score was greater than one in all basins indicating an over-forecast of rainfall events. Besides, the performance of model forecasts slightly degraded with the increase in the lead time of forecast.

A multi-category verification of forecasting rainfall events showed that the model under-forecasted the frequency of 'Heavy Rain' and 'No Rain' categories in most of the sub-basins, whereas it over-forecasted 'Light Rain' and 'Moderate Rain' categories. Continuous verification of 3-day temperature forecasts for four stations showed very good matching of forecasted temperatures with the observations as indicated by the PBIAS in the range of -3.1 to +3.23, and Skill Score in the range of 0.52 to 0.88. The WRF model slightly over-forecasted mean daily temperature in stations 417 (Rani Jaruwa) and 515 (Ghorahi), whereas it was slightly under-forecasted in stations 508 (Tulsipur) and 511 (Salyan Bazaar).

A multi-category verification of 3-day temperature forecasts showed mixed results with category 1 (cold condition) under-forecasted in station 417 (Rani Jaruwa) and over-forecasted in stations 508 (Tulsipur) and 511 (Salyan Bazaar). The category 2 (warm condition) is under-forecasted in station 511 (Salyan Bazaar) and over-forecasted in stations 508 (Tulsipur) and 515 (Ghorahi). The category 3 (hot condition) is over-forecasted in station 511 (Salyan Bazaar) and under-forecasted in stations 508 (Tulsipur) and 515 (Ghorahi).

We evaluated the performances of eight bias correction schemes for all 3-days rainfall forecasts for all sub-basins. It is found that the percent biases of rainfall forecasts were reduced significantly by all bias correction schemes but empirical quantile scheme has reduced the biases to a minimum. However, other schemes like robust empirical quantile, linear and exponential asymptotic schemes also reduced the biases significantly. Hence, empirical quantile method is the best bias correction scheme to correct biases in WRF model rainfall forecast for the sub-basins of the BRB. However, in case of parametric transformations exponential asymptotic function can also be considered as it has also reduced the biases in rainfall forecast significantly.

Evaluation of the performances of eight bias correction schemes for 3-day temperature forecasts for all four stations showed the percent bias of less than 5% by all bias correction schemes. However, empirical quantile, robust empirical quantile, smoothing spline, linear and exponential asymptotic schemes have reduced the bias significantly. In general, WRF model is quite capable of forecasting mean temperature with less than 5% bias. Hence, bias correction may not be necessary for mean temperature forecast. However, linear transformation, robust empirical quantile, empirical quantile, and smoothing spline could be considered as appropriate bias correction schemes to correct biases in the mean temperature for the stations of the BRB if deemed necessary.

Acknowledgments

This research has been funded by the Government of India, Ministry of Earth Sciences (Sanction Order No. MoES/RIMES/4/09(Pt.) under the project 'Technical Assistance for Development of Flood Forecast Generation and Application System for Disaster Mitigation in Nepal'. We would like to thank Nepal Department of Hydrology and Meteorology and Regional Integrated Multi-Hazard Early Warning System for Africa and Asia for providing necessary data to conduct this research.

References

Bannister, D., Orr, A., Jain, S. K., Holman, I. P., Momblanch, A., Phillips, T., Adeloye, A. J., Snapir, B., Waine, T. W., Hosking, J. S., and Allen-Sader, C., 2019. Bias Correction of High-Resolution Regional Climate Model Precipitation Output Gives the Best Estimates of Precipitation in Himalayan Catchments. *Journal of Geophysical Research: Atmospheres*, 124, 14 220–14 239, doi: 10.1029/2019jd030804.

DHM, 2017. Observed climate trend analysis in the districts and physiographic regions of Nepal (1971-2014). *Department of Hydrology and Meteorology, Kathmandu*, 1, 74.

Dudhia, J., 2004. The weather research and forecasting model (version 2.0). *Proceedings of the 2nd International Workshop on Next Generation NWP Model*, Yonsei University, 19–23.

Fowler, T. L., Jensen, T. L., and Brown, B. G., 2012. Introduction to Forecast Verification. *University Corporation for Atmospheric Research*.

Gautam, D., and Regmi, S., 2013. Recent trends in the onset and withdrawal of summer monsoon over Nepal. *Ecopersia*, 1 (4), 353–367.

Gibilisco, S., 2005. Meteorology Demystified. *McGraw Hill Professional*.

Gilleland, E., 2010. Confidence intervals for forecast verification. NCAR Technical Note. Tech. rep., NCAR/TN-479+ STR. http://citeseerx. ist. psu. edu/viewdoc/download.

Gudmundsson, L., 2016. qmap: Statistical transformations for post-processing climate model output. R package version 1.0-4.

Gudmundsson, L., Bremnes, J. B., Haugen, J. E., and Engen-Skaugen, T., 2012. Downscaling RCM precipitation to the station scale using statistical transformations–a comparison of methods. *Hydrology and Earth System Sciences*, 16 (9), 3383–3390.

Paulhus, J. L., and Kohler, M. A., 1952. Interpolation of missing precipitation records. *Monthly Weather Review*, 80 (8), 129–133.

R Core Team, 2013. R: A language and environment for statistical computing. Vienna, Austria, R Foundation for Statistical Computing, URL https://www.R-project.org/.

Shrestha, M. S., Gurung, M. B., Khadgi, V. R., Wagle, N., Banarjee, S., Sherchan, U., Parajuli, B., and Mishra, A., 2021. The last mile: Flood risk communication for better preparedness in Nepal. *International journal of disaster risk reduction*, 56, 102 118.

Yates, E., Anquetin, S., Ducrocq, V., Creutin, J.-D., Ricard, D., and Chancibault, K., 2006. Point and areal validation of forecast precipitation fields. *Meteorological Applications*, 13 (1), 1–20.

Young, K. C., 1992. A three-way model for interpolating for monthly precipitation values. *Monthly Weather Review*, 120 (11), 2561–2569.