WEATHER RESEARCH AND FORECASTING MODEL (WRF)

VERIFICATION

2020

Weather and Climate Services Division

National Center for Hydrology and Meteorology

2020

Table of Contents

1. Introduction

National Center for Hydrology and Meteorology (NCHM) currently runs Environmental Modeling System Weather Research and Forecasting (EMSWRF) for the daily weather forecasting with lead time of 3 days (72 hours). The EMSWRF has been installed and operational since November 2015 with the support from the Finnish Meteorological Institute (FMI) under the project of Strengthening Hydromet Services for Bhutan. The EMSWRF is a Local Area Model which is used for downscaling of the weather forecast information to a finer grid using the boundary condition from the Global Model with coarser grid information. Bhutan being a mountainous country has varying weather and climate within a short distance.

The information generated from EMSWRF is used as a guidance by the regular forecaster to produce the daily weather forecast. As such there is a need to understand the performance of the weather model to use the information effectively.

NCHM produced an initial Weather Research and Forecasting (WRF) Model Verification Report in 2019. Therefore, in continuation to the report, additional 15 stations with data of 2019 shall be validated, using simple statistical methods to compare the performance of EMSWRF with the observation data from the field.

2. Data and Methodology

2.1 Observation data

Meteorological variables of surface temperature and rainfall is used for the verification of the WRF model. There are 20 Agrometeorological stations (Class A) across the country which are identified as the focal point of weather forecasting for Bhutan. The Class A stations are manned by NCHM staff and report data to the National Weather and Flood Warning Center (NWFWC) in Thimphu twice a day at 9:00 AM and 3:00 PM BST. The basic weather variables such as temperature (maximum and minimum), 24 hours accumulated rainfall and cloud octas are reported. Besides these variables, Class A station also measures other weather variables.

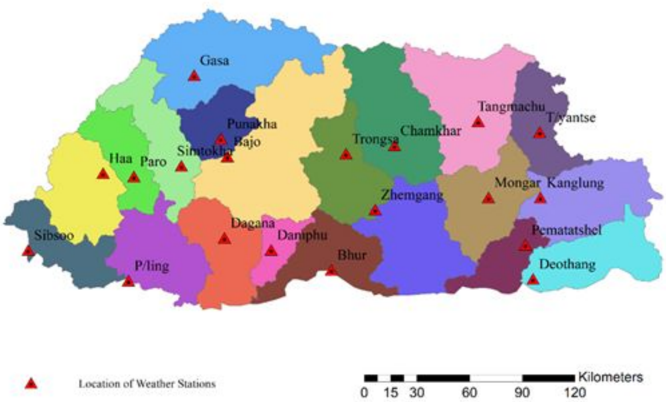

Figure 1: Location of 20 Agrometeorological stations (Class A)

For this report analysis, observation data from the 20 Class A stations are compared with the forecast data of WRF from year 2016 till 2019 (1 January 2016 till 31 December 2019).

2.2 Model output

The model used for the analysis is Environmental Modeling System (EMS) version 3.4 which is a complete, full-physics, state-of-the science numerical weather prediction (NWP) package that incorporates both the NOAA(NEMS) and Weather Research and Forecasting (WRF) model system into a single user-friendly, end-to-end forecasting system. All the capability of the National Centers for Environmental Prediction (NCEP) NEMS and National Center for Atmospheric Research (NCAR) WRF models are retained within the EMS. Nearly every element of an operational NWP system has been integrated into the EMS, including the acquisition and processing of initialization data, model execution, output data processing, and file migration and archiving. Even tools for the display of output are provided. Real-time forecasting operations are enhanced through the use of an automated process that incorporates various failover options as well as the synchronous post processing and distribution of forecast files. The EMS can run on either a stand-alone workstation or a cluster of Linux computers.

EMSWRF V3.4 Model runs every 6 hours for initial conditions of 00, 06, 12 and 18 UTC. The model has the capacity to run with a lead time of 72 hours (3days). The model runs with a nested domain of 45 vertical levels with the parent domain and nested domain with the horizontal resolution of 15 km and 3 km respectively (Fig.1_The location of domains and their sizes). The boundary initial conditions used for the model is from the Global Forecast System (GFS) model, NCEP, NOAA, which is a coupled model (atmosphere, ocean, land/soil and sea ice) with 64 vertical levels and has a horizontal resolution of 28 km (details-attached in the annexure). The WRF data period used for the analysis is January 2016 to December 2018, 12 UTC run.

Dynamics	Non-Hydrostatics
Model Domain	1. Outer Domain: Lon: 79° 27’ 16.26” E, 102° 51’ 18.25” E Lat: 16° 48’ 02.59” N, 37° 00’ 36.70” N 2. Inner Domain: Lon: 84° 58’ 28.39” E, 95° 54’ 54.59” E Lat: 23° 01’ 14.63” N, 31° 03’ 48.82” N
Primary Time step	67
Vertical Layers	45
Grid Spacing	Outer Domain (15km) Inner Domain (3 km)
Map Projection	Lambert (2016- April 2018), Mercator since April 2018
Radiation parameterization	Ra_sw_physics : Dudhia Scheme Ra_lw_physics : RRTM Scheme (Mlawer et al.,1997,JGR)
Cumulus scheme	Kain-Fritsch Scheme (Kain, 2004, JAM) for outer domain. Default for inner domain.
PBL Scheme	YSU Scheme (Hong, Noh and Dudhia,2006,MWR)
Microphysics scheme	Lin (Purdue) Scheme (Lin, Farley and Orville, 1983, JCAM)

Table 1: Details of WRF model set up

Figure 2: Location of domains and their sizes

2.3 Continuous Variable Analysis

Verifying forecasts of continuous variables measures how the values of the forecasts differ from the observations. Verification of continuous forecasts often includes exploratory plots such as scatter plots and box plots, as well as various summary scores. The scatter plots give a first look at correspondence between forecast and observations. An accurate forecast will have points on or near the diagonal. The box plots boxes to show the range of data falling between the 25th and 75th percentiles, horizontal line inside the box showing the median value, and the whiskers showing the complete range of the data. It shows similarity between location, spread, and skewness of forecast and observed distributions. However, it does not give information on the correspondence between the forecasts and observations.

Following are the statistical analysis done for the report;

i. Standard Deviation (SD)
ii. Mean Error (ME) or Bias
iii. Mean Absolute Error (MAE)
iv. Root Mean Square Error (RMSE)

2.4 Dichotomous variable analysis

We defined the event before creating a dichotomous variable. Defining the event- according to the World Meteorological Organization (WMO, 2014), it says that the nature of the event must be predicted and must be clearly stated in order to understand what is being predicted and the location. Accordingly, the contingency table for rainfall is prepared refer Table 2) with model run as 'Event Forecast (yes/no)' and observed station rain 'Event Observed (yes/no)' to collect a match set of forecast and observation. Rainy day is termed when a station and model records 1 mm or more rainfall in a day as per Indian Meteorological Department (IMD).

		Event observed		Marginal total
		Yes	**No**	
Event	**Yes**	A	B	A+B
Forecast	**No**	C	D	C+D
Marginal total		A+C	B+D	A+B+C+D

(A=Hit, B=False alarm, C=Miss, D= Correct Rejection)

Table 2: Contingency table for dichotomous variable analysis

- '**Hit**' is defined by occurrence of at least one observation of rainfall anytime during the forecast valid time.
- '**False alarm**' is defined when rainfall is forecast, but there is no rainfall observed in the forecast area.
- '**Miss**' is an actual record of rainfall during the valid day, but it was not reported in the forecast.
- '**Correct rejection**' is when there is no forecast of rainfall and there was no record of rainfall reported on the valid day.

2.4.1 Calculating scores using the contingency table

From the contingency table generated from model and observation data for the rainfall, following scores are computed to get the result of analysis;

i. Frequency bias (B)
ii. Probability of detection (PoD)
iii. False alarm ratio (FAR)
iv. Post Agreement (PAG)
v. False alarm rate (F)
vi. Hanssen-Kuipers score (KSS)
vii. Heidke skill score (HSS)
viii. True skill statistic (TSS)

3. Analysis and Results

3.1 Continuous Variables

The analysis of the continuous variable is represented in time series, scatter plots and box plots for the forecast days (Day 1, Day 2 and Day 3) for all the 20 Class A stations. Mean, Median, Standard Deviation (SD), Mean Error (ME), Mean Absolute Error (MAE), Root Mean Square Error (RMSE) as well as a correlation has been calculated and tabulated.

3.1.1 Maximum Temperature Analysis

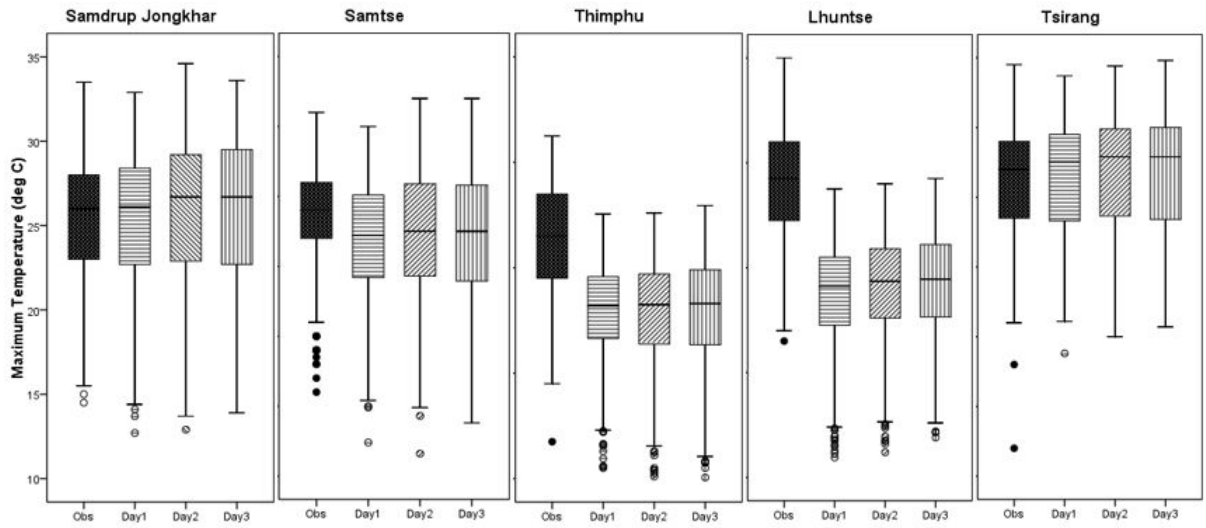

Figure 3: Boxplot of maximum temperature for forecast Day 1, Day 2 and Day 3 for 5 selected stations

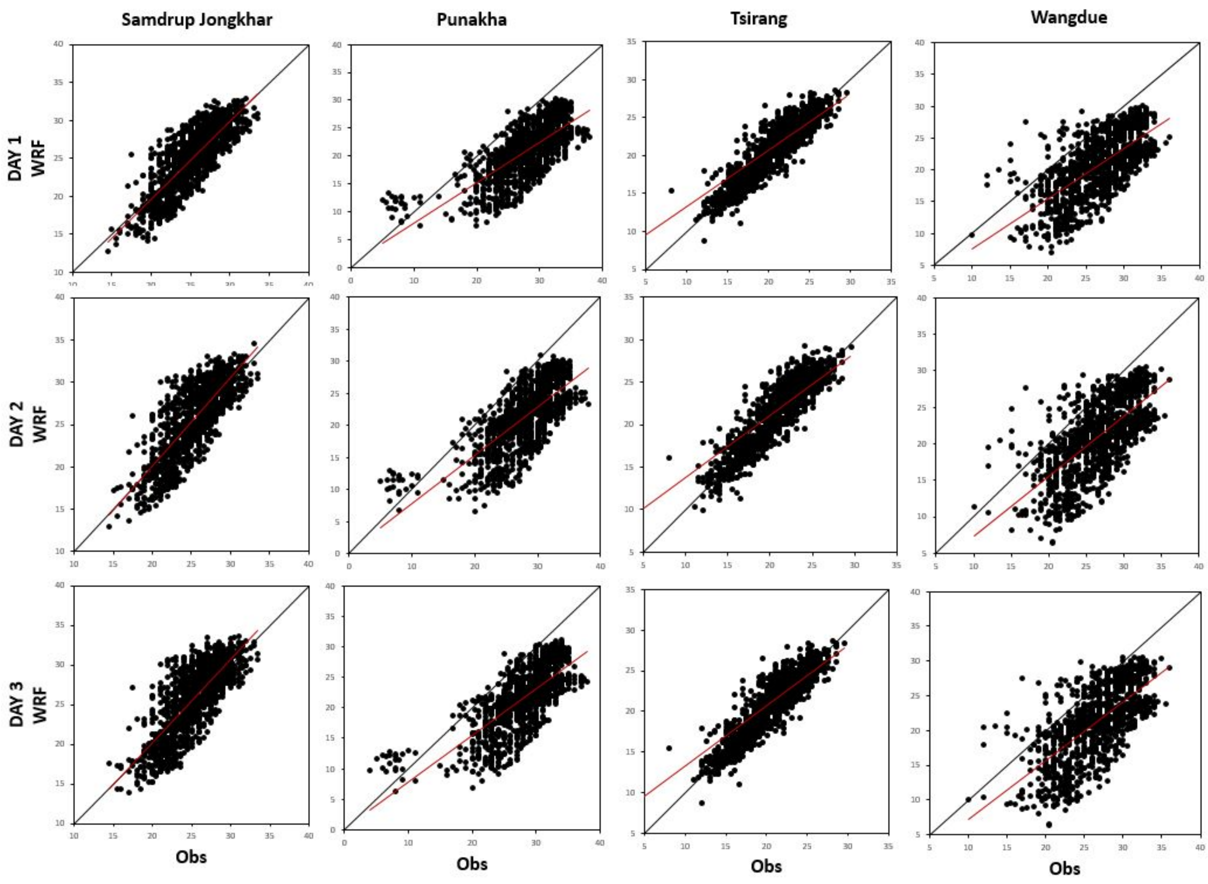

Figure 4: Scatter plot for maximum temperature for forecast Day 1, 2 and 3 for 4 selected stations

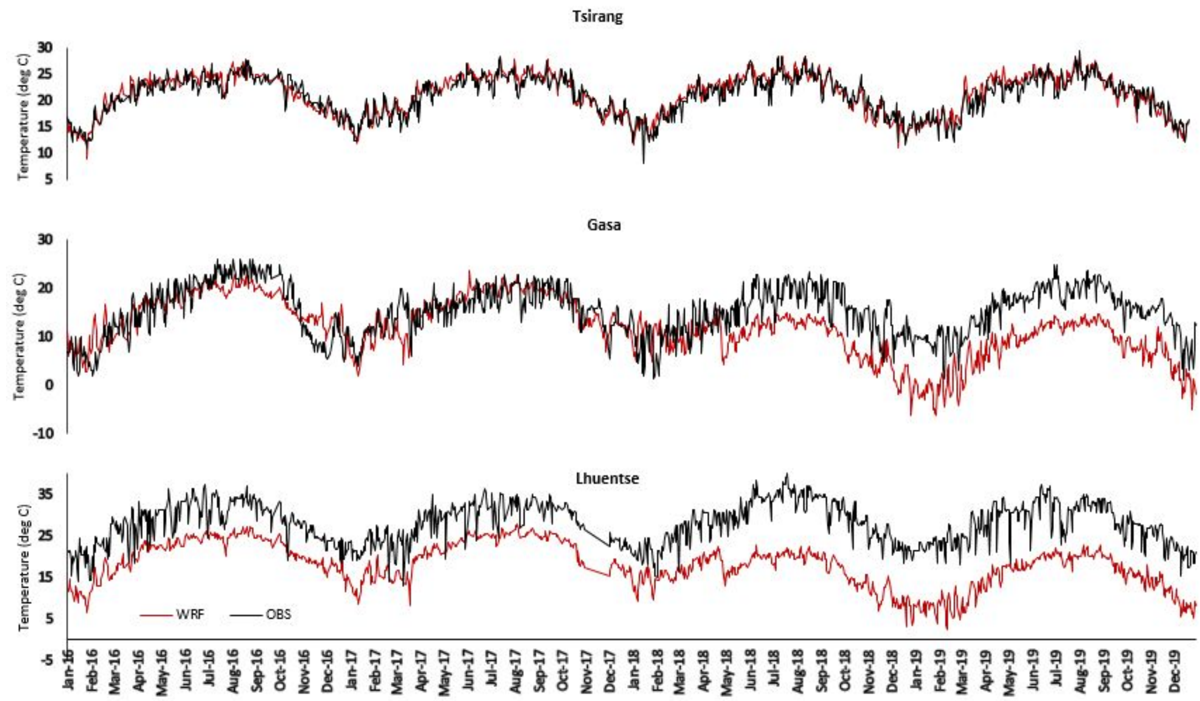

Figure 5: Times series plot for maximum temperature for forecast Day 1 for 3 selected stations

The box plot analysis for maximum temperature shows increasing dispersion of values of forecast from Day 1 to 3 to observation. The observation dispersion of values is lesser than the forecast days for Phuentsholing, Samtse and Sarpang. The maximum temperature prediction was relatively well for Samdrup Jongkhar and Tsirang. Rest all other stations were underpredicted. Highest underprediction can be seen in Lhuntse with an average of -10 °C. Larger outliers are indicated in Gasa, Haa, Lhuentse, Paro Punakha and Thimphu.

The scatter plot analysis shows that the maximum temperature for all the forecast days, Day 1,2 and 3 for all the stations presented similar dispersion. The analysis shows positive linear correlation association for all forecast days between the maximum temperature forecast and observed for all the stations. Clustered values along the line of best fit with few noticeable dispersed values which illustrates that most of the forecast was captured with few uncaptured extreme values were seen at most of the stations for all 3 days. However, Dagana, Lhuentse, Tashi Yangtse, Paro, Trongsa, Bumthang and Wangdue showed larger dispersion for all 3 days. The prediction was well for Samdrup Jongkhar for all 3 days. However, it was underpredicted for the rest of the stations except for Tsirang throughout the period. Outliers can be seen mostly at Punakha and Wangdue.

The time series analysis shows that the maximum temperature for all the forecast days, Day 1,2 and 3 for all the stations presented similar trends. The maximum temperature was relatively well for Mongar, Samdrup Jongkhar, Phuentsholing, Zhemgang, Samtse, Pema Gatshel, Sarpang and Tsirang stations for all the four years from 2016-2019. However, for Gasa, Haa, Paro, Bumthang and Dagana stations it was started underpredicting from early summer season in 2018. Rest of the stations were under predicted for all the years. The highest underprediction can be seen in Lhuentse with an average of -10 °C.

Days		Mean	Medium	SD	ME	MAE	RMSE	Correlation
				Samdrup Jongkhar				
Obs		25.4	25.5	3.3				
WRF	Day 1	25.1	25.9	4.0	-0.2	5.4	23	0.8
	Day 2	25.6	26.5	4.3	0.3	7.1	2.7	0.8
	Day 3	25.8	26.7	4.4	0.4	7.9	2.8	0.8
				Zhemgang				
Obs		20.4	21.0	4.4				
WRF	Day 1	18.8	19.2	4.3	-1.6	9.9	3.1	0.8
	Day 2	19.2	19.8	4.5	-1.3	9.7	3.1	0.8
	Day 3	19.4	20.0	4.5	-1.1	9.5	3.1	0.8
				Tsirang				
Obs		20.9	21.5	3.7				
WRF	Day 1	21.2	22.2	3.8	0.5	6.8	2.6	0.9
	Day 2	21.7	22.8	3.9	0.9	8.4	2.9	0.9
	Day 3	21.8	23.0	4.0	1.1	8.4	2.9	0.9
				Sarpang				
Obs		27.5	27.5	3.1				
WRF	Day 1	25.3	26.1	4.0	-2.2	13.7	3.7	0.7
	Day 2	25.8	26.7	4.2	-1.7	13.4	3.7	0.7
	Day 3	25.9	26.8	4.3	-1.6	1.5	3.7	0.7
				Pema Gatshel				
Obs		22.0	22.0	4.1				
WRF	Day 1	20.4	21.2	3.8	-1.6	7.7	2.8	0.8
	Day 2	20.7	21.7	3.9	-1.2	7.4	2.7	0.8
	Day 3	18.5	18.7	0.9	-3.4	26.7	5.2	0.3

Table 3: Analysis table for maximum temperature for selected 5 stations

For maximum temperature the bias was negligible for Samdrup Jongkhar for all the forecast days, Day 1, 2 and 3 and for Tsirang for day 1. It was underpredicted for all the stations for all forecast days ranging from -0.2 for day 1 for Samdrup Jongkhar to -10.6 for day 1 for Lhuentse. However, there was over prediction of value 0.3 and 0.4 values for day 2 and 3 respectively for Samdrup Jongkhar and for Tsirang for all 3 days with average bias of 0.8. The mean error was observed to be decreasing from day 1 to 3 for most of the stations except for Pema Gatshel (increase in bias on 3rd day), and for Samdrup Jongkhar and Tsirang (increases with lead time).

The RMSE remained the same for Sarpang (3.7), Phuentsholing (3.6), Zhemgang (3.1) and Samtse (3.5) for all forecast days. RMSE for Tsirang was the same for day 2 and 3 (2.9). Samdrup Jongkhar and Pema Gatshel showed an increase in RMSE on the 3rd day. Rest of the stations showed decrease with lead time.

3.1.2 Minimum Temperature Analysis

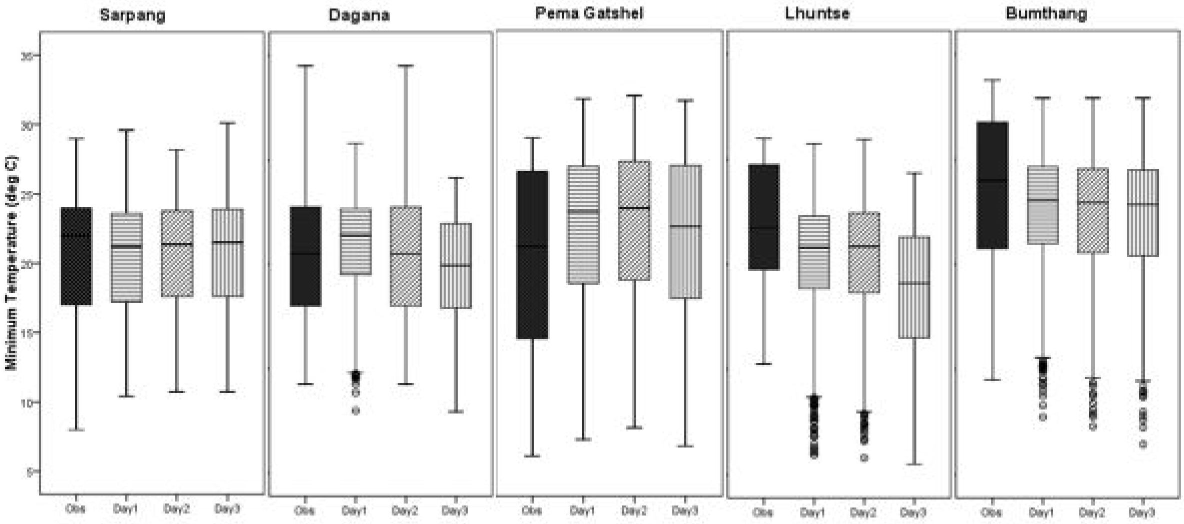

Figure 6: Boxplot of Minimum Temperature for 5 stations with forecast days (Day 1, Day 2 and Day 3)

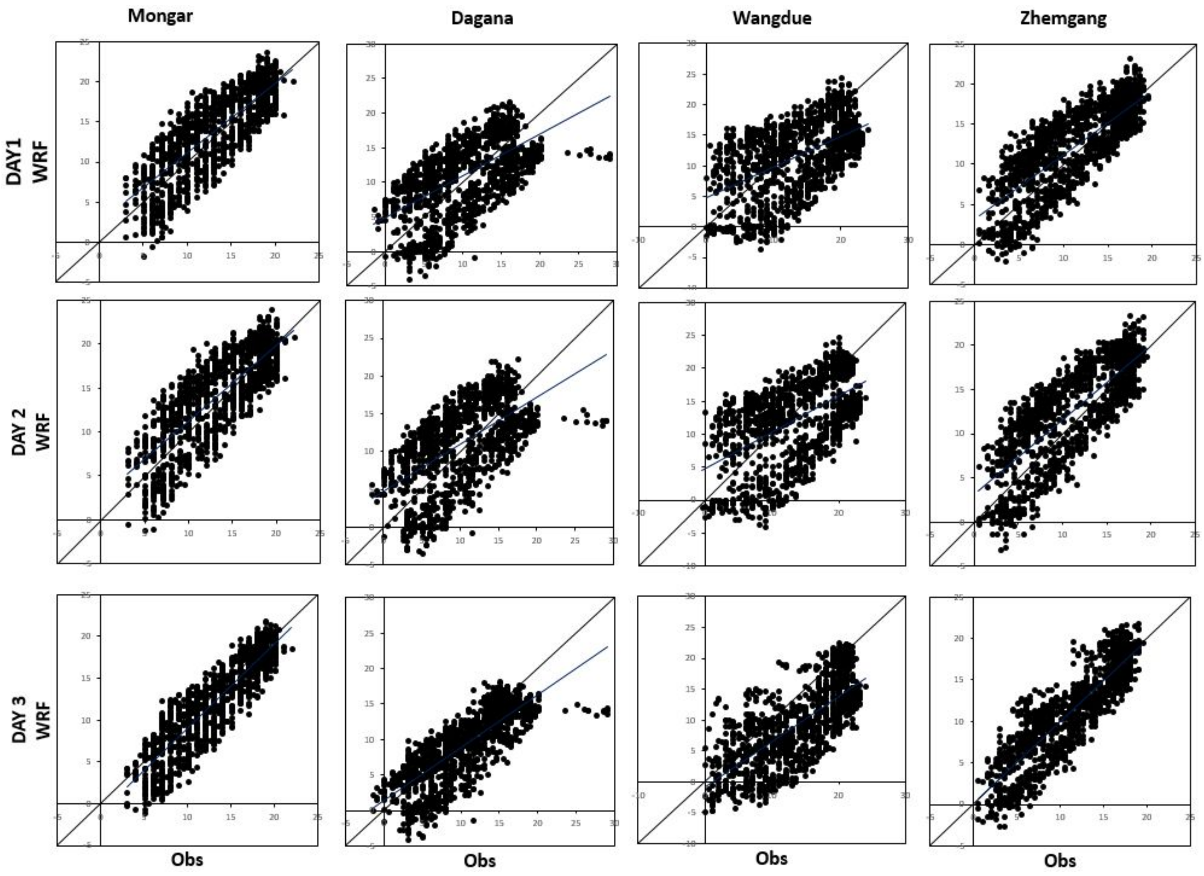

Figure 7: Scatter plot for Minimum Temperature for 3 stations with forecast days (Day 1, Day 2 and Day 3)

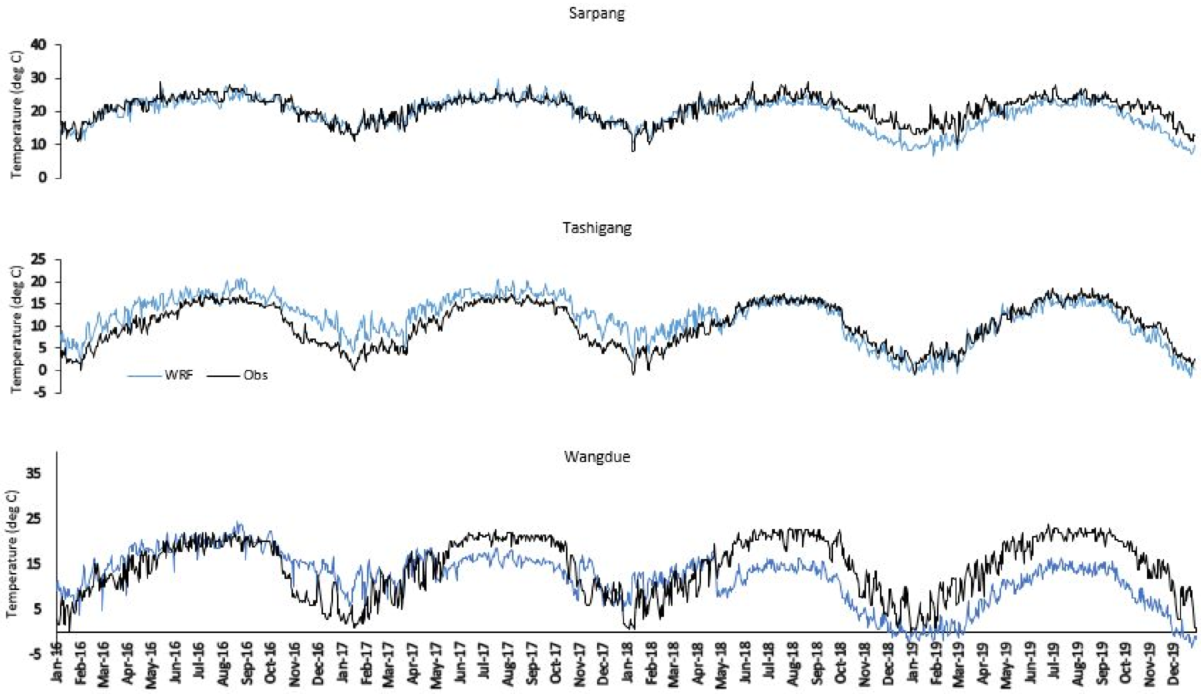

Figure 8: Times series plot for minimum temperature for forecast Day 1 for 3 selected stations

The box plot analysis for minimum temperature shows similar dispersion of observation with forecast of Day 3 for most of the stations except for Phuentsholing where the observation dispersion is lesser than the forecast dispersion, whereas it is the opposite for Pema Gatshel, Haa and Tsirang. The minimum temperature prediction was relatively well for Sarpang for all 3 days. The predictions for day 1 and 2 were relatively good for most of the stations except for Lhuentse, Punakha and Bumthang where it was underpredicted. The forecast for day 3 was also more or less captured except for Haa, Lhuntse, Paro, Punakha, Tashi Yangtse and Lhuentse where it was underpredicted. Outliers can be seen mostly in Gasa, Lhuentse, Bumthang and Paro.

The scatter plot analysis shows that the minimum temperature for all the forecast days, Day 1,2 and 3 for all the stations presented similar dispersion. The analysis shows linear correlation association for all forecast days between the minimum temperature forecast and observed for all the stations. Clustered values along the line of best fit with few noticeable dispersed values which illustrates that most of the forecast was captured with few uncaptured extreme values were seen at most of the stations except few such stations such as Lhuentse, Phuentsholing, Tashi Yangtse, Gasa, Haa, Paro and Wangude which showed larger dispersions for all 3 days. The prediction was relatively well for Mongar. It was underpredicted mostly for Lhuentse, Tashi Yangtse, Punakha, and Sarpang. Rest of the stations were overpredicted Outliers can be seen mostly in Dagana.

The time series analysis shows that the minimum temperature for forecast days, Day 1, 2 and 3 represented similar trends. The minimum temperature was relatively good for Lhuentse, Haa, Trongsa, Sarpang and Tashi Yangtse till early summer season (April, May) in 2018, and thereafter for Mongar, Phuentsholing, Samdrup Jongkhar, Samtse, Pema Gatshel, Tsirang and Tashigang. Dagana, Gasa, Haa, Mongar and Paro showed similar results where the minimum temperature was over predicted till early summer 2018 and thereafter under predicted. The minimum temperature was underpredicted from early summer 2017 for Punakha, Tashi Yangtse, Trongsa and Wangdue.

		Mean	Medium	SD	ME	MAE	RMSE	Correlation
				Gasa				
Obs		4.8	5.0	6.4				
WRF	Day 1	4.9	6.1	6.4	0.0	29.4	5.4	0.6
	Day 2	4.8	6.2	6.6	0.0	30.1	5.5	0.6
	Day 3	0.4	0.8	5.8	-4.4	29.0	5.4	0.9
				Lhuentse				
Obs		13.7	13.5	5.5				
WRF	Day 1	10.1	11.3	6.2	-3.6	35.0	5.9	0.7
	Day 2	10.2	11.5	6.2	-3.4	34.2	5.8	0.7
	Day 3	7.6	8.2	6.0	-6.1	46.9	6.8	0.9
				Phuentsholing				
Obs		18.5	19.5	4.3				
WRF	Day 1	20.5	21.8	4.5	2.0	14.1	3.8	0.7
	Day 2	20.6	22.0	4.	2.1	15.1	3.9	0.7
	Day 3	18.5	19.1	4.5	-0.1	8.5	2.9	0.8
				Bumthang				
Obs		6.6	7.5	7.1				
WRF	Day 1	4.7	5.7	5.8	-1.9	28.9	5.4	0.7
	Day 2	4.7	5.8	6.0	-2.0	29.6	5.4	0.7
	Day 3	4.6	5.	6.1	-2.0	30.3	5.5	0.7
				Punakha				
Obs		15.5	16.0	6.3				
WRF	Day 1	11.7	13.1	5.8	-3.7	45.9	6.8	0.6
	Day 2	12.3	13.5	6.2	-3.2	44.0	6.6	0.6
	Day 3	9.7	10.0	6.4	-5.8	55.6	7.5	0.7

Table 4: Analysis table for minimum temperature for selected 5 stations

For minimum temperature the bias was negligible for Gasa for the first and second day and for Zhemgang, Samtse and Phuentsholing on the third day. It was underpredicted for Lhuentse, Tashi Yangtse, Trongsa, Wangdue, Punakha, Bumthang , Thimphu and Sarpang for all 3 days with bias ranging from-6.1 for Lhuentse on day 3 to -0.2 for Thimphu on day 1. Haa, Dagana, Gasa, Samtse, Mongar, Paro, Phuentsholing showed underprediction on the 3rd day. Rest of the stations showed over prediction with bias ranging from 0.1 for Haa on the 2nd day to 2.9 for Samdrup Jongkhar on the same day. Most of the stations showed an increase in bias from day 1 to day 2, followed by decrease in bias on the 3rd day. However, Haa and Paro showed decrease in error with lead time.

The RMSE remained the same for Haa (6.6), Bumthang (5.4), Trongsa (4.8) and Thimphu (5.2) for day 1 and 2. It also remained the same for Gasa (5.4) for day 1 and 3 and Tsirang (3.3) for day 1 and 3. Lhuentse and Punakha showed decrease in RMSE from day 1 to 2 followed by increase on day 3. However it was the reverse with the rest of the stations.

3.2 Dichotomous Variables Analysis

After sorting the rainfall events from the respective stations for the model and observation data, the contingency table is generated and further scores are computed for each station and forecast days (Day 1, Day 2 and Day 3).

3.2.1 Categorical predictands analysis

Stations	B	POD	FAR	PAG	F	KSS	HSS	TS
				Day 1				
Thimphu	1.2	0.5	0.6	0.4	0.2	0.3	0.3	0.3
Samdrup Jongkhar	2.7	1.0	0.6	0.4	1.0	0.0	0.0	0.4
Samtse	0.8	0.6	0.2	0.8	0.1	0.5	0.5	0.6
Paro	1.2	0.6	0.5	0.5	0.2	0.4	0.4	0.3
Punakha	0.9	0.5	0.5	0.5	0.1	0.3	0.4	0.4
				Day 2				
Thimphu	1.1	0.5	0.6	0.4	0.2	0.3	0.3	0.3
Samdrup Jongkhar	0.7	0.5	0.2	0.8	0.1	0.4	0.5	0.5
Samtse	0.8	0.7	0.2	0.8	0.1	0.6	0.6	0.6
Paro	1.2	0.6	0.5	0.5	0.2	0.4	0.4	0.3
Punakha	0.9	0.5	0.5	0.5	0.1	0.3	0.4	0.3
				Day 3				
Thimphu	1.0	0.5	0.5	0.5	0.2	0.3	0.3	0.3
Samdrup Jongkhar	0.7	0.5	0.3	0.4	0.1	0.4	0.4	0.4
Samtse	0.8	0.6	0.2	0.8	0.1	0.5	0.5	0.5
Paro	1.2	0.6	0.5	0.5	0.2	0.4	0.4	0.3
Punakha	0.9	0.5	0.5	0.5	0.2	0.3	0.3	0.3

Table 5: Computed scores for Rainfall using contingency table (annexure)

The Frequency Bias Index (B) ranges up to value 1.3 for Dagana for all 3 days,Tashigang for day 2 and 3 and for Tsirang for day 2. The index for Dagana (1.3), Paro (1.2), Punakha (0.9), Tashi Yangtse and Zhemgang (1.1), Samtse and Sarpang (0.8) remain the same over the 3 days. There was a decrease in index from day 1 to 2, followed by an increase on the 3rd day for Lhuentse, Mongar, Trongsa, Wangdue and Pema Gatshel. Thimphu showed a decrease in index.

The Probability of Detection (POD) or Hit Rate remained almost constant for most of the stations for all 3 days, with values 0.4-0.7, meaning the forecast was able to capture the event of rainfall 40-70% respectively. Gasa indicated 0.4 POD for day 1 and 3 and Samdrup Jongkhar showed 1.0 for day 2.

The False Alarm Ratio (FAR) and its inverse Post agreement (PAG) remained almost the same for most of the stations with FAR value ranging from 0.2 for Samtse to 0.5 for Dagana, Mongar, Paro, Punakha and Tashigang indicating 20% and 50% decrease respectively in the forecast of rainfall event captured. Lhuentse and Wangdue showed an increase in FAR by 0.1 value on the 3rd day.

False alarm rate (F) ranges from 0 to 1 and 0 being a perfect score (WMO, 2014). It illustrates that the model has made a forecast of rainfall for the valid period but it didn't occur during the valid period. All the stations show F between 0.1- 0.2 which means that 10% to 20% of the forecast was False Alarm. The model performs well for the Sarpang station with only 10% of False alarm rate.

The False alarm rate (F) for Phuentsholing, Tashi Yangtse, Pema Gatshel, Sarpang and Tashigang showed an increase in rate of 0.1 from day 2 to 3. However, Samdrup Jongkhar and Trongsa showed decrease in rate. Samdrup Jongkhar indicated the highest F of 1.0 on the 2nd day. The F remained the same for most of the stations with rates 0.1 to 0.2 for all 3 forecast days, indicating 10 and 20 % increase in the forecast unable to capture the event of rainfall.

Hanssen & Kuiper`s skill (KSS) score and Heidke Skill Score (HSS) ranges from 0.3 to 0.5 for most of the stations indicating 30% and 50% decrease in the forecast unable to capture the rainfall event respectively, for all 3 days. Samdrup Jongkhar represented the highest decline in KSS and HSS from 0.4 to 0.0 from day 2 to 3.

Most of the stations showed the same Threat score (TS) over the 3 days with score range from 0.3 to 0.5, indicating a 30 to 50% forecast was able to capture the rainfall event. The lowest score of 0.3 was at Paro, Punakha and Thimphu for the whole 3 days. Samtse indicated the highest threat score with 0.6 for day 1 and 3.

3.2.2 Histogram

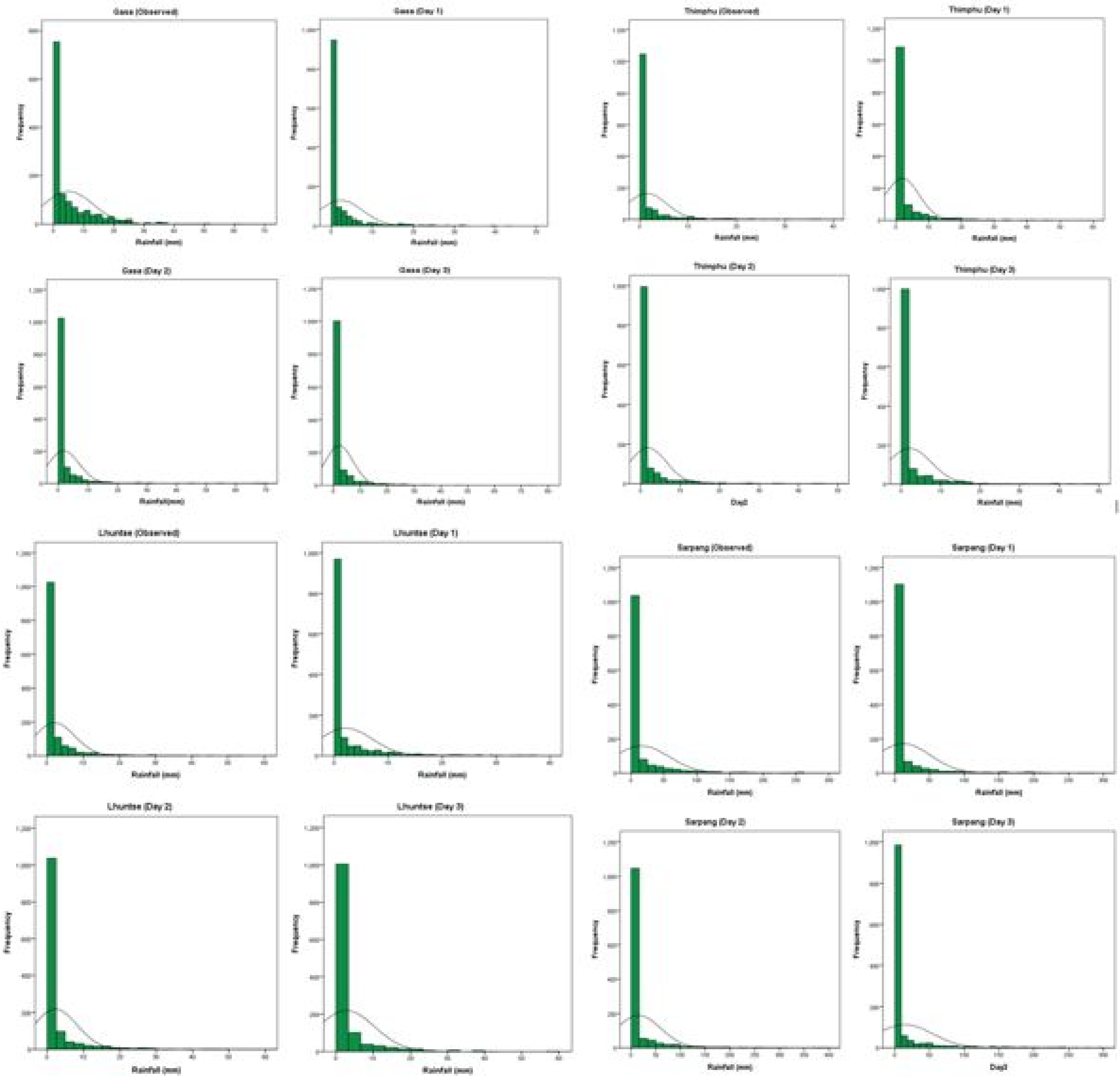

Figure 9: Histogram for Gasa, Thimphu, Lhuntse and Sarpang

The curve line tail tends to extend out to the right which is illustrated to be rightly skewed for most of the stations. The frequency of rainfall recorded at most of the stations was from 0-20 mm. The highest rainfall frequency was at Sarpang with value 0-50 mm.

Thimphu recorded the lowest extreme rainfall of 44.0mm while model extremes for day 1, 2 and 3 were 71.7 mm, 46.2 mm and 59.1 mm respectively. The highest frequency of rainfall for model and observation recorded was between 0-15 mm for model and observation. Gasa and Lhuentse showed a similar extreme rainfall record of around 60 mm with model extremes for day 1 and 2 between 37-82 mm whereas day 3 indicated a lower value of 61.5 mm for Gasa and Lhuentse showed 106.2 mm.

The highest extreme observation rainfall and forecast for day 1, 2 and 3 were recorded at Sarpang. It showed a value of 445.0 mm extreme observation rainfall. Forecast extreme average of 431.5 m was indicated from day 1 to 3.

4. Monthly Temperature and Rainfall Average Analysis

The monthly average temperature and rainfall between the observation and forecast from 2016-2019 were plotted to understand whether the WRF model was able to capture the monthly variation of temperature and rainfall for the country.

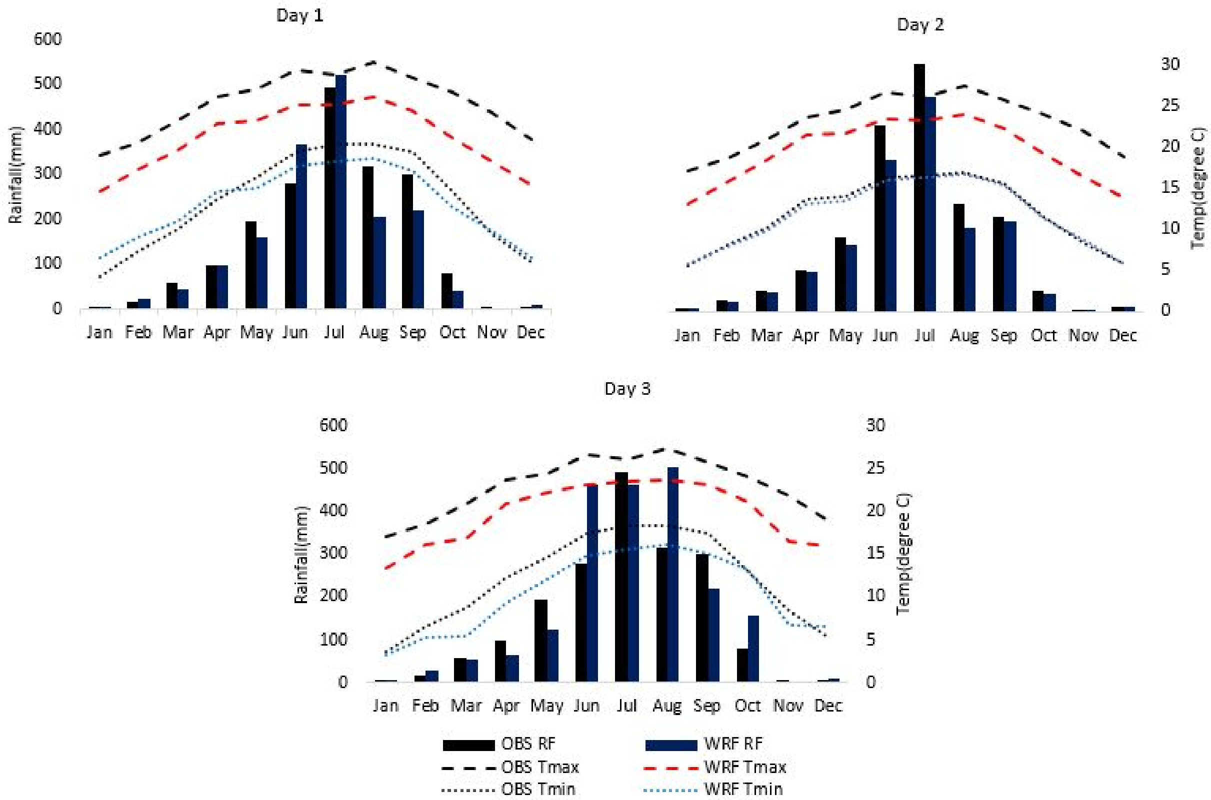

Figure 10:Monthly Average for model and observation

From the monthly average temperature and rainfall graphs, it can be seen that the WRF model was more or less able to capture the monthly variation of both maximum and minimum temperature, and rainfall. The maximum temperature for all 3 days was under predicted with bias of 4-5 °C. However, the minimum temperature for day 1 was over predicted from January till April and henceforth under predicted till November and December where the prediction values were almost similar. The minimum temperature forecast for day 2 was mostly captured. It was under predicted for day 3.

The monthly average rainfall was mostly under predicted for all 3 days with minimum bias, except for day 3's June and August where it was largely over predicted with value almost to 200 mm. The rainfall trend was more or less captured for seasons such as winter (December, January, February), spring (March, April, May) and autumn (October, November). The peak summer season in July was also mostly captured with bias of 29.7 mm, -90.0mm and -28.5 mm for day 1, 2 and 3.

	Maximum Temperature			Minimum Temperature			Rainfall		
	Obs	**WRF**	**Bias**	**Obs**	**WRF**	**Bias**	**Obs**	**WRF**	**Bias**
Jan	17.1	13.0	-4.1	3.6	5.7	2.1	5.5	7.3	1.8
Feb	18.6	15.6	-3.0	6.5	8.0	1.5	16.7	23.5	6.9
Mar	21.0	17.7	-3.3	8.9	9.8	0.9	58.9	46.1	-12.8
Apr	23.7	20.6	-3.0	12.2	13.1	0.9	97.6	96.9	-0.7
May	24.5	20.9	-3.6	14.7	13.4	-1.3	194.5	161.0	-33.5
Jun	26.6	22.8	-3.9	17.4	15.9	-1.6	277.8	366.8	89.0
Jul	26.1	22.8	-3.3	18.4	16.4	-2.1	492.2	521.9	29.7
Aug	27.4	23.6	-3.9	18.4	16.7	-1.7	316.6	204.7	-111.9
Sep	25.7	22.0	-3.7	17.4	15.4	-2.0	300.1	219.4	-80.7
Oct	24.1	19.0	-5.1	13.1	11.4	-1.7	80.7	41.1	-39.6
Nov	21.8	16.5	-5.3	8.4	8.7	0.3	5.6	3.2	-2.4
Dec	18.8	13.9	-5.0	5.3	5.7	0.4	5.5	10.2	4.6
Avg	22.9	19.0	-3.9	12.0	11.7	-0.4	154.3	141.8	-12.5

Table 8: Monthly average for model and observation for Day 1

5. Conclusion

Validation of Weather Research and Forecasting model (WRF) for variables of surface maximum and minimum temperature and rainfall for the year 2016 and 2019 (1 January 2016 to 31 December 2019) by comparing the forecast data with the observation data from 20 Agrometeorological stations using simple statistical methods has been done.

The WRF model forecast for surface maximum temperature was relatively well for Mongar, Samdrup Jongkhar, Phuentsholing, Zhemgang, Samtse, Pema Gatshel, Sarpang and Tsirang for all 3 days. However, for Gasa, Haa, Paro, Bumthang and Dagana stations it was started under predicting from early summer season in 2018. Rest of the stations were under predicted for all the days. The highest under prediction can be seen in Lhuentse with an average of -10 °C. The RMSE remained the same for Sarpang, Phuentsholing, Zhemgang and Samtse for all 3 days. Pema Gatshel showed an increase in RMSE on the 3rd day. Rest of the stations showed decrease in RMSE with lead time.

The model forecast for surface minimum temperature was relatively well for Sarpang, Mongar for all 3 days. The predictions for day 1 and 2 were relatively good for most of the stations except for Lhuentse, Punakha, Tashi Yangtse and Bumthang where it was under predicted. The forecast for day 3 was also more or less captured except for Haa, Lhuentse, Paro, Punakha, Tashi Yangtse and Lhuentse where it was under predicted. Rest of the stations were over predicted. The RMSE remained the same for Haa, Bumthang, Trongsa, Thimphu, Gasa and Tsirang for 2 days. Lhuentse and Punakha showed decrease in RMSE from day 1 to 2 followed by increase on day 3. However, it was the reverse with the rest of the stations.

Rainfall events were captured well with a bias of ± 0.2 for all the stations, however, Samdrup Jongkhar showed a greater bias of 2.7 for day 1. PAG, FAR, POD and F were found to be good for most of the stations located at the western and southern part of the country except Samdrup Jongkhar.

The WRF model was able to capture the monthly variation of both maximum and minimum temperature, and rainfall. The maximum temperature for all 3 days was under predicted with bias of 4-5 °C. However, the minimum temperature for day 1 was over predicted from January till April and henceforth under predicted. The minimum temperature forecast for day 2 was mostly captured. It was under predicted for day 3.

The monthly average rainfall was mostly under predicted for all 3 days with minimum bias, except for day 3's June and August where it was largely over predicted with value almost to 200 mm. The rainfall trend was more or less captured for seasons such as winter (December, January, February), spring (March, April, May) and autumn (October, November).

6. References

- World Meteorological Organization. (2014). *Forecast Verification for African Severe Weather Forecasting: Demonstration Projects. Geneva, Switzerland: World Meteorological Organization.*
- A Nearly Complete Guide to the UEMS V15
- Weather Research and Forecasting, ARW Version 3 Modeling System User's Guide, July 2015. *Mesoscale and Microscale Meteorology Division, National Center for Atmospheric Research*.
- Verification on NWP products _ Lecture for NWP training course. *WMO Regional Training Centre, Nanjing, China*
- Verification Techniques and Simple Theoretical Forecast Models. Robert Fawcett *National Climate Centre, Australian Bureau of Meteorology, Melbourne, Victoria, Australia*
- WRF Model Verification, *National Center for Hydrology and Meteorology, 2019*

Annexure-A

List of parameters observed and recorded from Class A stations;

1. Temperature
2. Rainfall
3. Relative humidity
4. Sunshine hours
5. Evaporations
6. Soil temperature at different depths (5 m, 15 m, 30 m)
7. Wind direction
8. Wind speed

Annexure-B

Methodology

1. Continuous variables

a. Standard Deviation (SD)

It is a measure of the amount of variation (or deviation) that might be expected between the observed value and the forecast value. It is a very concise and powerful way of conveying the amount of uncertainty in a forecast. The smaller the standard deviation, the less the uncertainty.

b. Mean Error (ME) Or Bias

It is the average error in a given set of forecasts. It represents a simple and informative score on the behavior of the given variable. If ME >0 (<0), the model exhibits over (under) forecasting. However it is not an accurate measure as it does not provide information of the mahniyure of errors. The value ranges from -∞ to +∞. The perfect score is equal to 0.

ME= (1/N) ∑ (fi-fo)

c. Mean Absolute Error (MAE)

It is the average magnitude of errors in a given set of forecasts. Therefore, it is a linear measure of accuracy. However, it does not distinguish between positive and negative forecast errors.
The value ranges from 0 to +∞. The perfect score is equal to 0.

MAE= (1/N) ∑ |fi-fo|

d. Root Mean Square Error (RMSE)

Measures "average" error, weighted according to the square of the error. Does not indicate the direction of the deviations. The *RMSE* puts greater influence on large errors than smaller errors, which may be a good thing if large errors are especially undesirable, but may also encourage conservative forecasting. The value ranges from 0 to +∞. The perfect score is equal to 0. *RMSE= (1/N) ∑ (fi-fo) ^2*

2. Dichotomous Variables: Calculating scores using the contingency table

a. Frequency Bias (B)

The frequency bias (B), it refers to as bias, uses only marginal sums of the contingency table. It compares the forecast and observed frequencies of occurrence of the event in the sample. The forecast is said to be unbiased if the event is forecast exactly the same frequency with which it is observed, so that the frequency bias of 1 represents the best score (WMO, 2014.

Frequency bias= a+b/ (a+c)

b. Probability of detection (PoD) (Hit rate (HR) or prefigurance)

The hit rate (HR) has a range of 0-1 with 1 representing a perfect forecast. It uses only the observed events a and c in the contingency table and it is sensitive only to missed events and not false alarms. The HR is incomplete by itself, so it is being used in conjunction with either false alarm ratio or false alarm rate as suggested in WMO demonstration project paper (2014).

PoD=HR=a/(a+c)

c. False alarm ratio (FAR)

The false alarm ratio (FAR) is the ratio of the total false alarms (b) to the total events forecast (a+b). It ranges from 0-1, 0 being a perfect score. It is insensitive to missed events. It is also incomplete score, so it should be used in connection with the HR [1] as suggested in WMO demonstration project paper (2014).

FAR= b/(a+b)

d. Threat score (TS) (Critical success index, CSI)

The threat score (TS), or critical success index (CSI), is frequently used as a standard verification measure. It is sensitive to hit, misses and false alarms. It ranges from 0-1, 1 being perfect score and 0 as no skill level. However, it is sensitive to climatological frequency of events (WMO, 2014).

CSI= a/(a+b+c)

e. The Heidke skill score (HSS)

Skill is the accuracy of a forecast compared with the accuracy of standard forecast. The HSS ranges from negative value to +1 (WMO, 2014).

HSS=2(ad-bc)/[(a+c) (c+d)+(a+b)(b+d)]

f. The false alarm rate (FA)

The false alarm rate is simply the fraction of observed non-events that are false alarms. As stated in the definition, false alarm rate is sensitive to false alarms only, not misses. The best score for the FA is 0. FA is used in connection with HR (Hit rate) in comparative sense (WMO, 2014).

FA=b/(b+d)

g. The Hanssen-Kuipers score (KSS) (Pierce score) (true skill statistic (TSS)

The Hanssen-Kuipers score (KSS) is also known as the true skill statistic (TSS). It is the difference between the hit rate and the false alarm rate. It measures the ability of the forecast to distinguish between occurrence and non-occurrence of the event. It ranges from -1 to 1, 1 being perfect score and 0 as no skill level (WMO, 2014).

KSS=TSS=POD-F

KSS=ad-bc/[(a+c) (b+d)]