རྒྱལ་ཡོངས་ཆུ་དཔྱད་དང་གནམ་གཤིས་རིག་པའི་ལྟེ་བ།
དཔལ་ལྡན་འབྲུག་གཞུང་།

WEATHER AND CLIMATE SERVICES DIVISION
NATIONAL CENTER FOR HYDROLOGY & METEOROLOGY
ROYAL GOVERNMENT OF BHUTAN

REPORT
ON

WRF VERIFICATION

2018-2019

Abstract

EMSWRF V3.4 Model 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. The three years analysis, using simple statistical methods of the WRF output (Temperature and Rainfall) with the observation from 5 focal point of the country has been done for the 3 days forecast.

The model has good performance for the maximum and minimum temperature forecast with some bias. Lower RMSE is shown over the Tsirang and Sarpang and high over the Thimphu, however, RMSE is found to be more or less improving for all the stations. We have seen a very good correlation between the model and observations ranging from 0.7 to 0.9 for Bhutan for temperature with better relation with southern Bhutan.

Rainfall events were captured well with the bias of ± 0.2 for all the stations, however, the bias was found to be increasing with the lead time. The model is able to capture the monthly variation for rainfall with the peak rainfall during July and drier over the winter season. The model over estimates the precipitation during rainy season (JJAS) with larger bias of 219mm for June and 200mm for July. All the stations have a score between 0.3-0.5 for KSS, HSS and TS. It illustrates that the model has accuracy of 30% to 50% for the rainfall event. The model underestimates the maximum temperature for all the months with average warmer bias of 3.3 °C, where the minimum temperature shows warmer bias during the summer and colder bias during the winter.

Key words. WRF, Temperature and Rainfall verification, mean, median, standard deviation, ME, RMSE.

Contents

1. Introduction

National Center for Hydrology and Meteorology (NCHM) currently runs EMSWRF for the daily weather forecasting with lead time of 3 days (72 hours). The EMSWRF is 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 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 short distance.

The information generated from EMSWRF are 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. Therefore, this report will try to validate, using simple statistical methods, the performance of EMSWRF with comparison to the observation data from the field.

2. Data and Methodology

2.1 Observation data

Meteorological parameter of surface temperature and rainfall is used for the verification of the WRF model. The total of 5 manual meteorological Class A (Agro met) stations are considered to represent the country and are also the focal point of forecasting identified for Bhutan. The manual meteorological station is manned by regular staff of NCHM and reports data to forecasting office two times every 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.

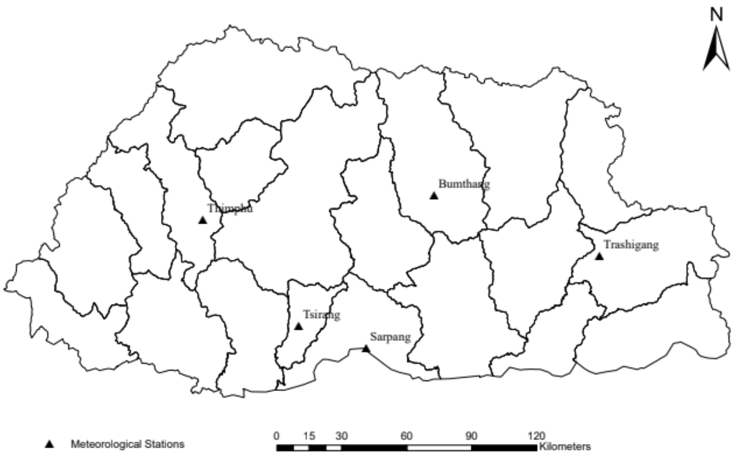

Figure 1: Location of Meteorological Station

`Table 1: Details of station location

Station Name	Latitude (N)	Longitude (E)	Elevation (m)
Bumthang	27.5403	90.7550	2470
Tsirang	27.0000	90.1217	1520
Thimphu	27.4383	89.6753	2310
Trashigang	27.2825	91.5222	1930
Sarpang	26.9039	90.4339	375

The observation data available from the manual station starts from 1996 till date and for this report analysis data period from 2016 is used since the weather model (WRF) run started from 2016 onwards.

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 condition of 00, 06, 12 and 18 UTC. The model has capacity to runs with 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.

Table 2: Details of WRF model setup

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)

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 plot 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 World Meteorological Organization (WMO, 2014), it says that the nature of event must being predicted 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).

Table 3: Contingency table for dichotomous variable analysis

		Event observed		Marginal total
		Yes	**No**	
Event Forecast	**Yes**	A	B	A+B
	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)

- '**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 actually there was a record of rainfall during the valid day, but it was not reported in forecast.
- '**Correct rejection**' is when there is no forecast of rainfall and there was no record of rainfall reported in 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 each of the five 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 (table 4).

3.1.1 Maximum Temperature Analysis

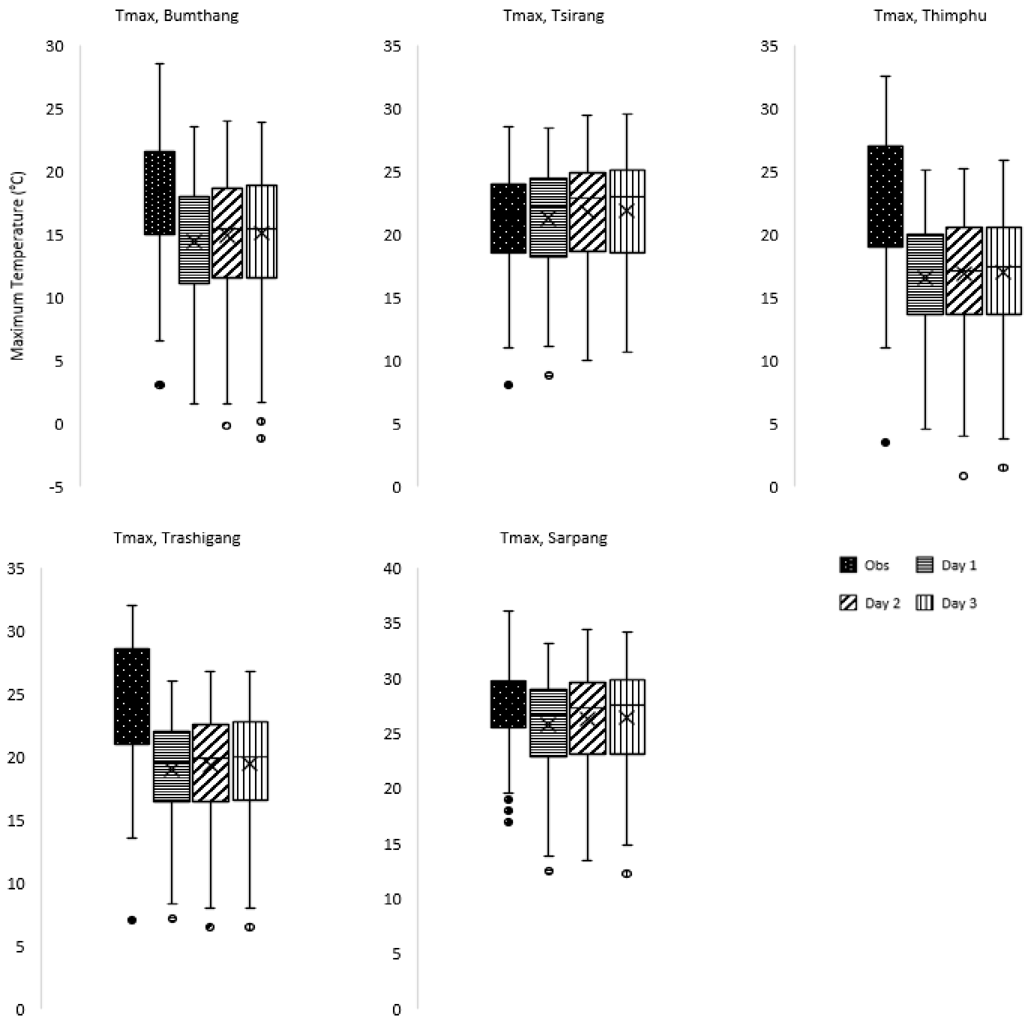

Figure 3: Boxplot of Maximum Temperature for 5 stations with forecast days (Day 1, Day 2 and Day 3

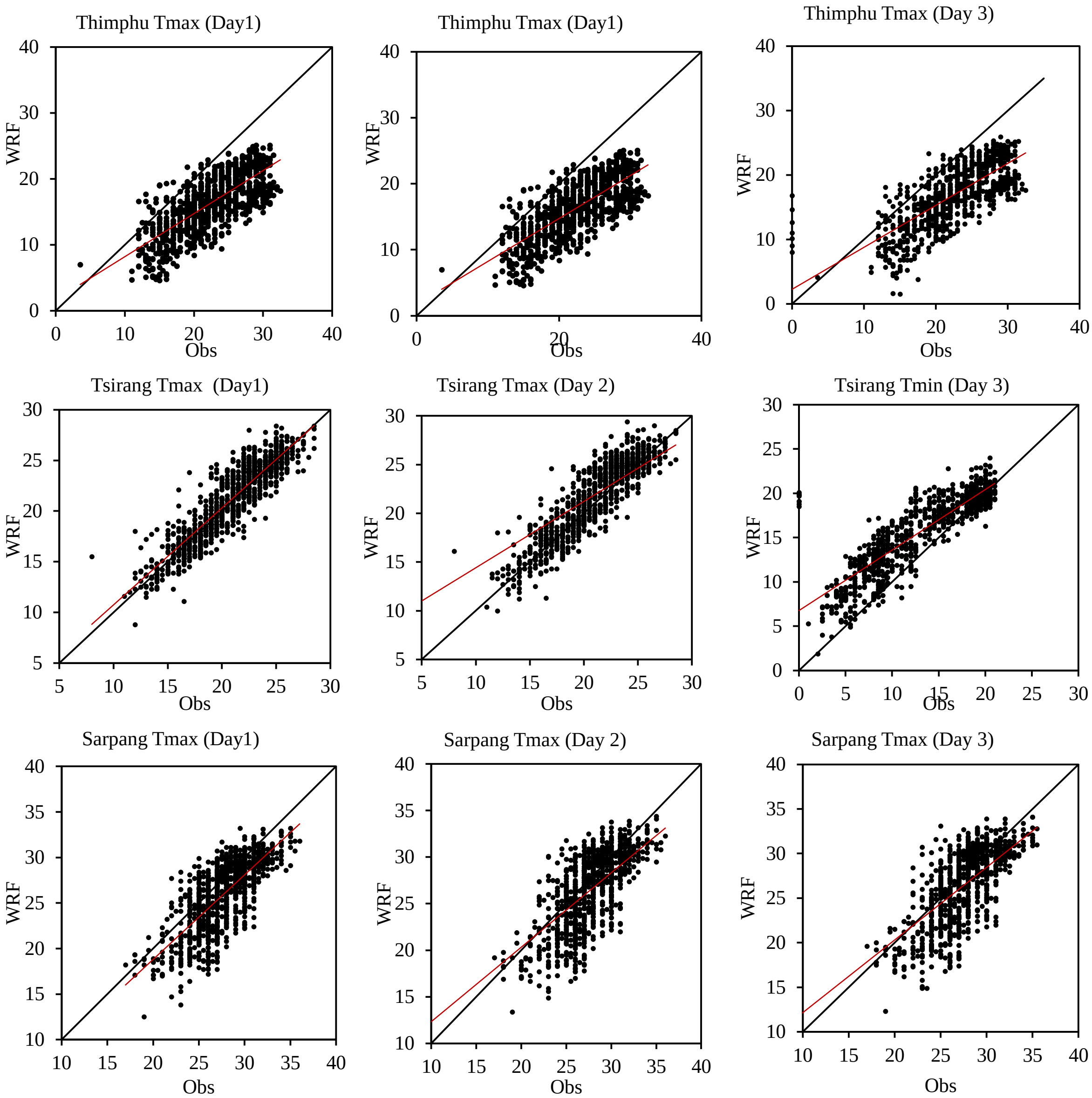

Figure 4: Scatter plot for Maximum Temperature for 3 stations with forecast days (Day 1, Day 2 and Day 3)

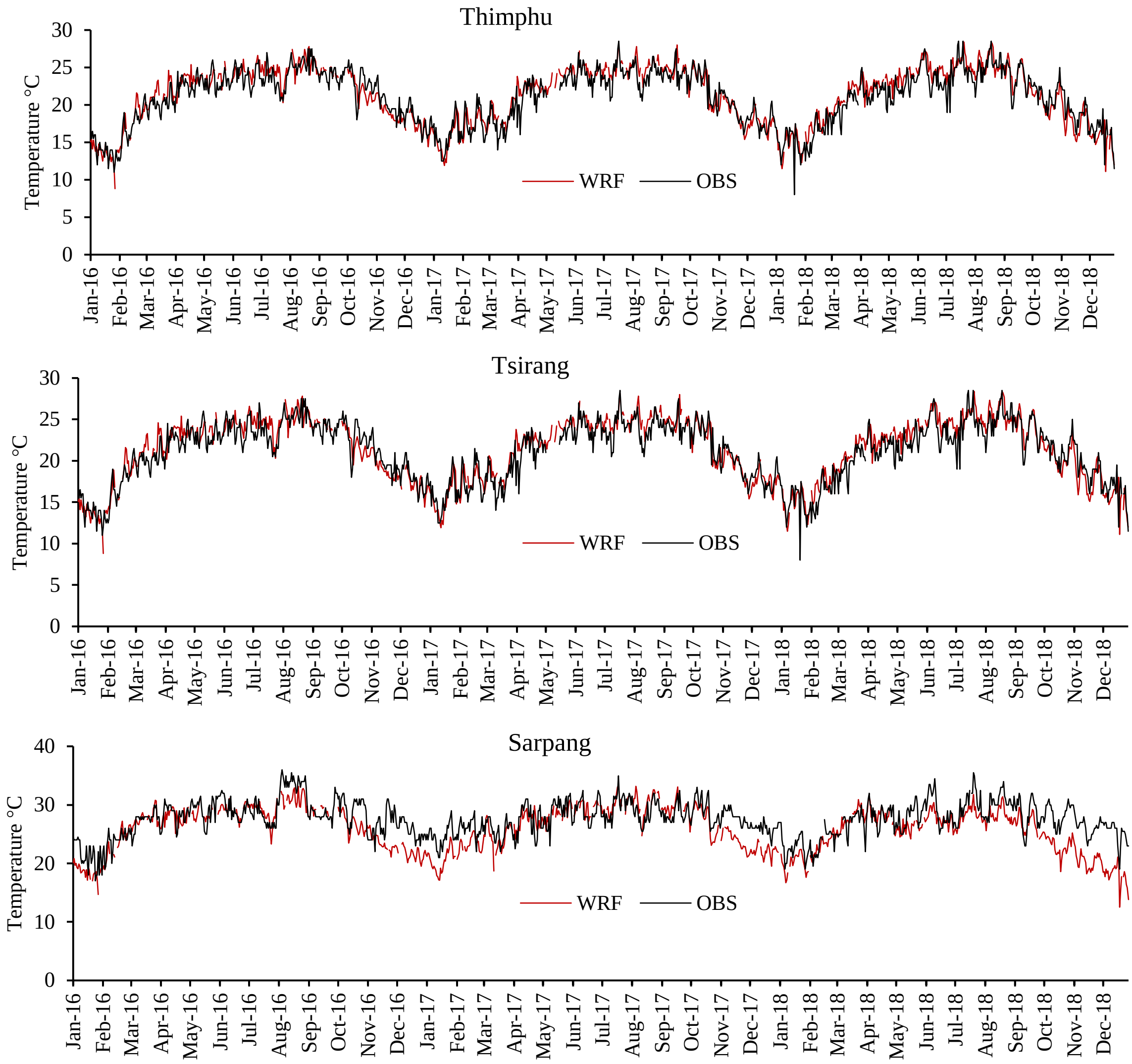

Figure 5: Times series plot for Maximum Temperature for 3 stations with forecast

From the boxplot (figure 3), it can be seen that, for Bumthang, the dispersion of values of the observation as well as forecast days are similar yet, it is underestimating the maximum temperature by 3.1 °C compared to their mean. For Tsirang and Sarpang, there are less dispersion of values than the rest of the forecast days. The model is over-estimating the maximum temperature by 0.6 °C for Tsirang and under-estimating the maximum temperature by 1.3 °C for Sarpang compared to their means. We can find that the model was able to forecast near to the observation.

There is more dispersion of values than rest of the forecast days in case of Trashigang and Thimphu. The model is under-estimating the maximum temperature by 5.3 °C for Trashigang while comparing their mean.

The scatter plot (figure 4) for Thimphu shows strong positive linear association between the maximum temperature forecasted and observed. For Trashigang and Sarpang, there are clustered values but there are noticeable dispersed values, which illustrates that the model has captured most of the observed maximum temperature, yet there are some extreme values unable to be captured by model. There are outliers in the maximum temperature.

The time series (figure 5) shows that the model over predicting for spring and summer seasons (March, April, May, June, July, August, September) and performing relatively well during the remaining months for Thimphu. The model is relatively performing well during the winter seasons for Trashigang. The model has relatively performed well for the maximum temperature during the spring and summer seasons but low performances during the winter seasons in case of Sarpang.

Table 4: Analysis table for Maximum Temperature

Days		Mean	Median	SD	ME	MAE	RMSE	Correlation
				Trashigang				
Obs		24.5	25.0	4.3				
WRF	Day 1	19.0	19.5	3.8	-5.5	35.4	6.0	0.9
	Day 2	19.4	19.9	4.0	-5.2	32.4	5.7	0.8
	Day 3	19.4	20.1	4.1	-5.0	31.0	5.6	0.8
				Sarpang				
Obs		27.5	28.0	3.1				
WRF	Day 1	25.8	26.6	3.8	-1.7	11.1	3.3	0.7
	Day 2	26.3	27.3	4.0	-1.2	11.2	3.3	0.7
	Day 3	26.4	27.5	4.2	-1.1	11.5	3.4	0.7
				Thimphu				
Obs		18.2	18.0	4.4				
WRF	Day 1	15.1	15.4	4.9	-3.1	21.1	4.6	0.8
	Day 2	16.8	17.1	4.6	-5.9	47.9	6.9	0.8
	Day 3	17.0	17.4	4.7	-5.6	44.9	6.7	0.8
				Tsirang				
Obs		21.0	21.5	3.6				
WRF	Day 1	21.3	22.2	3.8	0.5	7.8	2.8	0.9
	Day 2	21.8	22.9	3.9	1.0	9.6	3.1	0.9
	Day 3	15.1	15.4	4.9	1.1	9.4	4.6	0.9
				Bumthang				
Obs		18.3	18.5	4.3				
WRF	Day 1	14.5	14.9	4.7	-3.8	25.2	5.0	0.7
	Day 2	14.9	15.4	4.9	-3.4	22.4	4.7	0.7
	Day 3	15.1	15.4	4.9	-3.1	21.1	4.6	0.7

The model under predicts the maximum temperature for most of the stations with bias of about -5 °C for Trashigang, about -1°C for Sarpang and about -3°C for Thimphu and Bumthang.

Whereas for the Tsirang the model over predicts the maximum temperature with bias of above 0.5 °C. Overall, the model has a very good correlation with the observation ranging from 0.7 to 0.9 for Bhutan.

The model shows lower RMSE over the Tsirang and high over the Thimphu. The RMSE increases with lead time for Thimphu, Sarpang and Tsirang station whereas the Trashigang and Bumthang station shows reduced RMSE with lead time.

3.1.2 Minimum Temperature Analysis

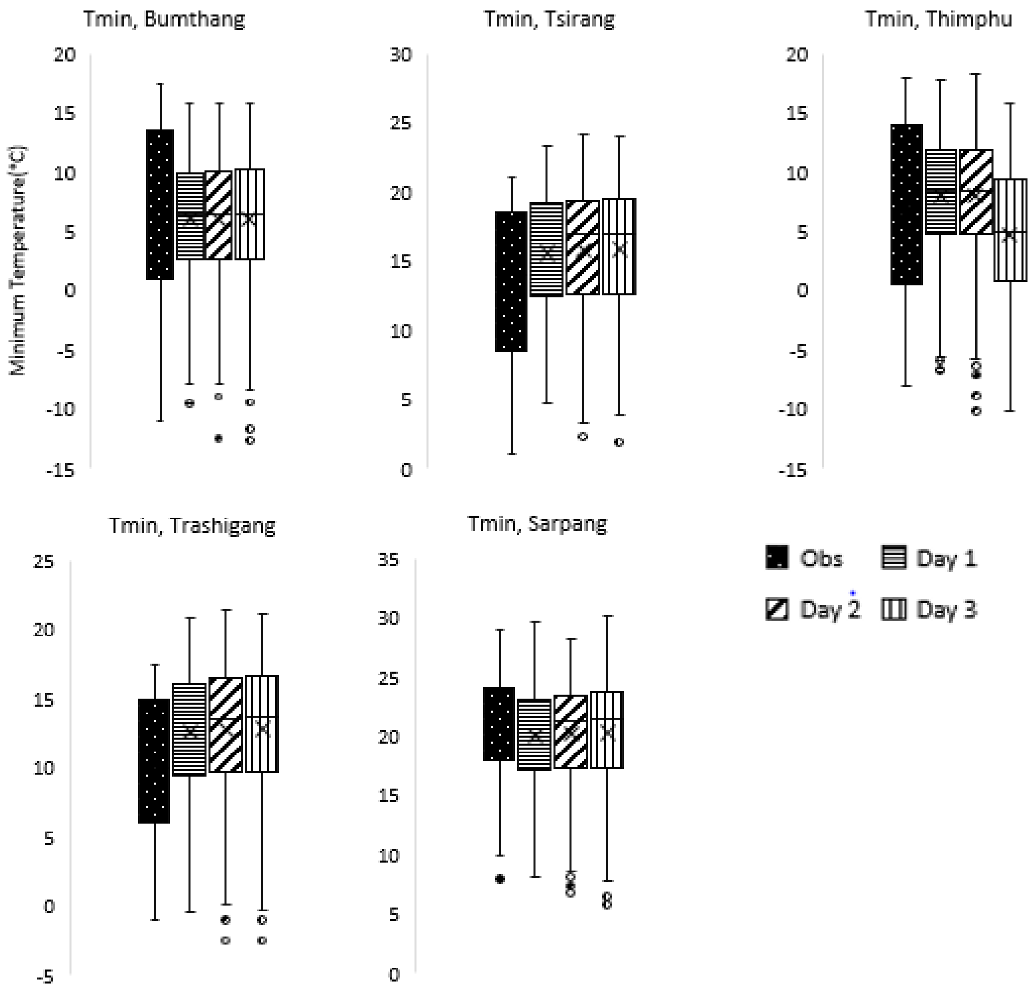

Figure 6: Boxplot of Minimum Temperature for 5 stations with forecast days (Day 1, Day 2 and Day 3)

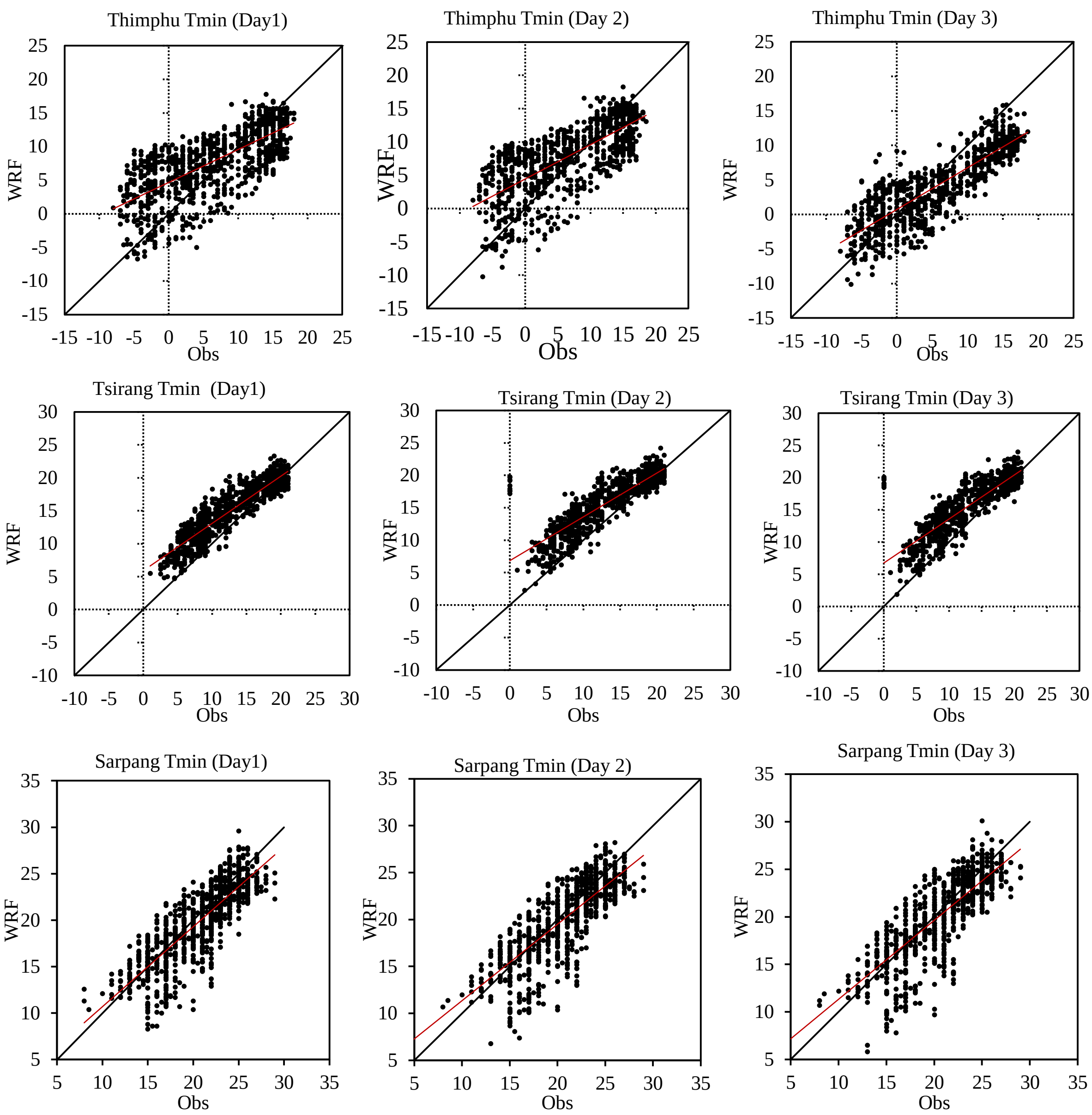

Figure 7: Scatter plot for Minimum Temperature for 3 stations with forecast days (Day 1, Day 2 and Day 3)

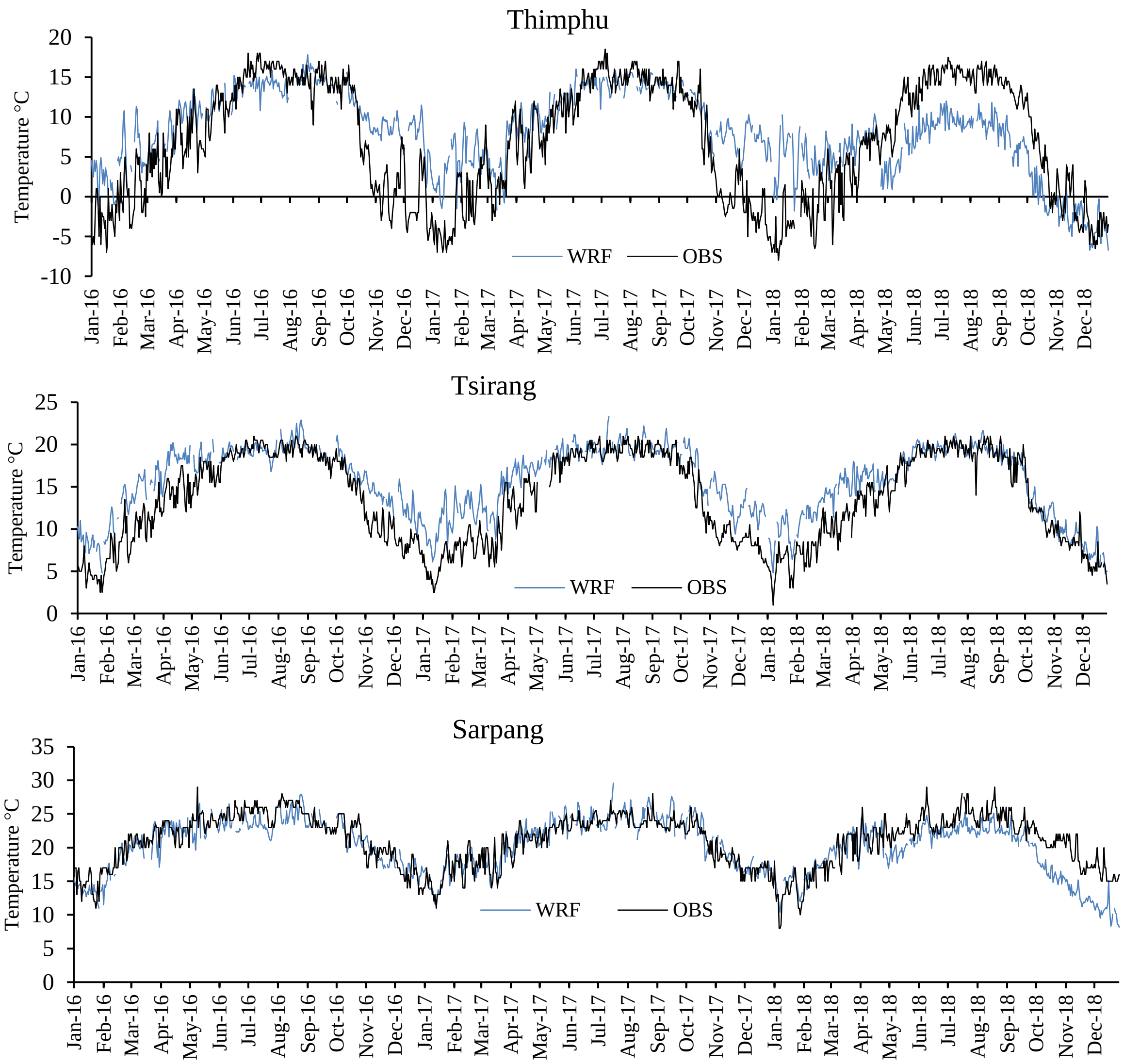

Figure 8: Time series plot for Minimum Temperature for 3 stations with forecast

The box plot (figure 6) illustrates much dispersion of values in the observed minimum temperature than rest of the forecast days for Thimphu, Trashigang, Tsirang and Bumthang. The model is over-estimating the minimum temperature by 0.4 ° C, by 2.5 ° C and by 2.3 ° C compared to their mean for first three stations and under-estimating the minimum temperature by 0.6 °C compared to their mean for Bumthang. The dispersion of values are similarly distributed for the observation as well as forecast days for Sarpang. The model is under-estimating the minimum temperature by 0.7 ° C while their means are compared.

The minimum temperature has very weak positive linear association for Thimphu (figure 7). In case of Trashigang and Sarpang, there is little dispersion of values, especially the extreme temperature, which illustrates that the model has failed to capture the extreme temperature. There are outliers in the minimum temperatures for all the stations.

The time series (figure 8) shows that it is over predicting for spring and summer seasons (March, April, May, June, July, August, September) and performing relatively well during the remaining months for Thimphu. The model has performed very well during the spring and summer seasons for the minimum temperature for Trashigang. The timeseries plot for Sarpang for all seasons is very good as it has well captured the season variability as well as the trend whereas towards the end of the series there is diversion due to possible change of model projection from Lambert to Mercator.

Table 5: Analysis table for Minimum Temperature

Days		**Mean**	**Median**	**SD**	**ME**	**MAE**	**RMSE**	**Correlation**
				Trashigang				
Obs		10.2	10.0	4.8				
WRF	Day 1	12.6	13.3	4.4	2.4	11.7	3.4	0.9
	Day 2	12.7	13.5	4.5	2.5	12.8	3.6	0.9
	Day 3	12.8	13.7	4.6	2.7	13.3	3.6	0.9
				Sarpang				
Obs		20.8	22.0	3.9				
WRF	Day 1	20.0	20.9	4.0	-0.8	6.8	2.6	0.8
	Day 2	20.2	21.3	4.1	-0.7	7.4	2.7	0.8
	Day 3	20.3	21.5	4.2	-0.5	7.3	2.7	0.8
				Thimphu				
Obs		6.6	7.0	7.2				
WRF	Day 1	6.0	6.5	5.3	-0.6	24.0	4.9	0.7
	Day 2	7.9	8.4	5.2	1.1	26.5	5.1	0.7
	Day 3	4.8	4.9	5.2	-1.9	19.4	4.4	0.9
				Tsirang				
Obs		13.4	14.0	5.3				
WRF	Day 1	15.6	16.5	4.1	2.3	11.9	3.5	0.9
	Day 2	15.8	17.0	4.2	2.5	13.8	3.8	0.9
	Day 3	6.0	6.5	5.3	2.6	14.2	4.9	0.9
				Bumthang				
Obs		6.6	7.0	7.1				
WRF	Day 1	6.0	6.4	5.0	-0.6	24.6	5.0	0.7
	Day 2	6.0	6.4	5.2	-0.6	24.3	4.9	0.7
	Day 3	6.0	6.5	5.3	-0.6	24.0	4.9	0.7

The model under predicts the minimum temperature with bias of about -0.5 °C for Sarpang, Thimphu and Bumthang whereas, it over predicts with bias of above 2.3 °C for Trashigang and Tsirang station. The RSME more or less improves for all the stations. The model shows very good correlation with the observation ranging from 0.7 to 0.9 for Bhutan.

3.2 Dichotomous Variables Analysis

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

3.1.1 Categorical predictands analysis

Table 6: Computed scores for Rainfall using contingency table (annexure)

Stations	B	POD	FAR	PAG	F	KSS	HSS	TS
				Day 1				
Bumthang	0.9	0.6	0.4	0.6	0.2	0.4	0.4	0.4
Tsirang	1.1	0.6	0.4	0.6	0.2	0.4	0.4	0.4
Thimphu	1.1	0.5	0.5	0.5	0.2	0.3	0.3	0.3
Trashigang	1.1	0.6	0.5	0.5	0.2	0.4	0.4	0.4
Sarpang	0.8	0.6	0.3	0.7	0.1	0.5	0.5	0.5
				Day 2				
Bumthang	0.8	0.5	0.4	0.6	0.2	0.3	0.3	0.3
Tsirang	1.3	0.7	0.5	0.5	0.2	0.4	0.4	0.4
Thimphu	1.1	0.5	0.5	0.5	0.2	0.3	0.3	0.3
Trashigang	1.2	0.6	0.5	0.5	0.2	0.4	0.4	0.4
Sarpang	0.9	0.7	0.3	0.7	0.2	0.5	0.5	0.5
				Day 3				
Bumthang	0.8	0.5	0.5	0.5	0.2	0.3	0.3	0.3
Tsirang	1.2	0.7	0.4	0.6	0.2	0.4	0.4	0.4
Thimphu	1	0.5	0.5	0.5	0.2	0.3	0.3	0.3
Trashigang	1.2	0.7	0.4	0.6	0.2	0.4	0.4	0.4
Sarpang	0.8	0.6	0.3	0.7	0.1	0.5	0.5	0.5

The frequency bias index (B) ranges from 0 to ∞ and 1 being perfect as stated in WMO (2014). The bias index for all the stations (table 6) are near to 1 with ± 0.2 variation showing little difference between forecast and observation. Most of the station shows increase in the frequency bias with increasing lead time.

Probability of Detection (POD), sometime called as Hit rate, ranges from 0 to 1, and 1 being the perfect score (WMO, 2014). POD for all the stations are between 0.5-0.7. It illustrates that, it has detected 50% to 70% of rainfall forecast occurred during the valid period of the forecast. We can see that stations located in the south detects more rainfall event than the stations located in the west and east, this is due to the fact that most of the rainfall event happens mostly during the summer season over the southern parts of the country.

False alarm ratio (FAR) ranges from 1 to 0 and 0 being perfect score (WMO, 2014). FAR for all the stations are between 0.3-0.5. It indicates that 30% to 50% of the forecast were not observed on the valid forecast period. Stations located in the south shows less FAR compared to the stations in the west, central and east.

Post agreement (PAG) ranges from 0 to 1 and 1 being perfect score. It illustrates that, the model has made a forecast of rainfall for the valid period and there was a rainfall during the valid period. The PAG for all stations are between 0.5-0.7 indicating 50% to 70% agrees the model forecast and the event occurred.

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.

Hanssen & Kuiper`s skill (KSS) score ranges from -1 to 1, where 1 is a perfect score and 0 is no skill level. Heidke skill score (HSS) ranges from $-\infty$ to 1, where 1 is the perfect score and 0 as no skill level. Threat score (TS) ranges from 0 to 1, where 1 as perfect score and 0 as no skill level (WMO, 2014). All the stations have a score between 0.3-0.5 for KSS, HSS and TS. It illustrates that the model has accuracy of 30% to 50% for the rainfall event.

3.1.1 Histograms

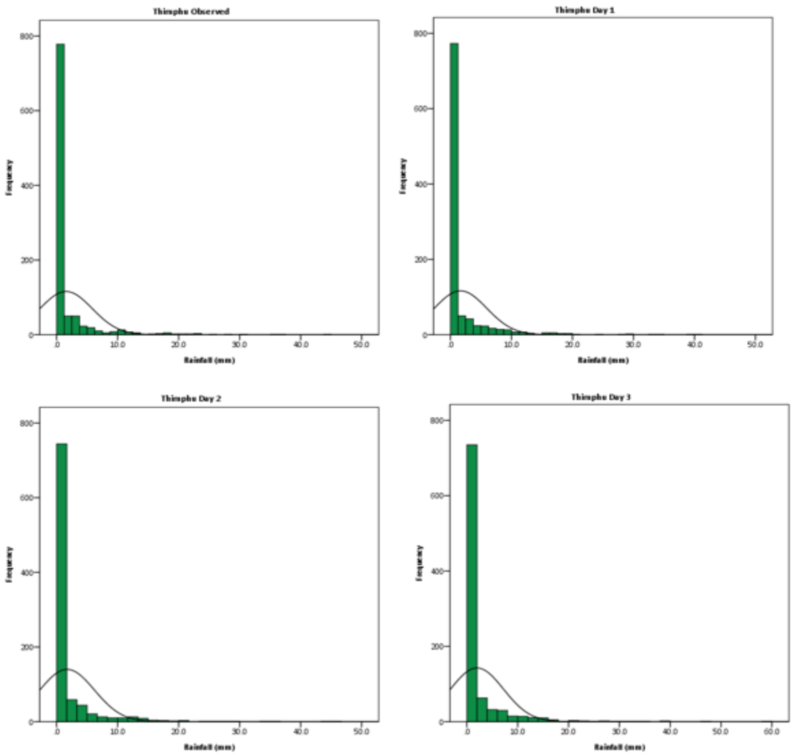

Figure 9: Histogram for Thimphu

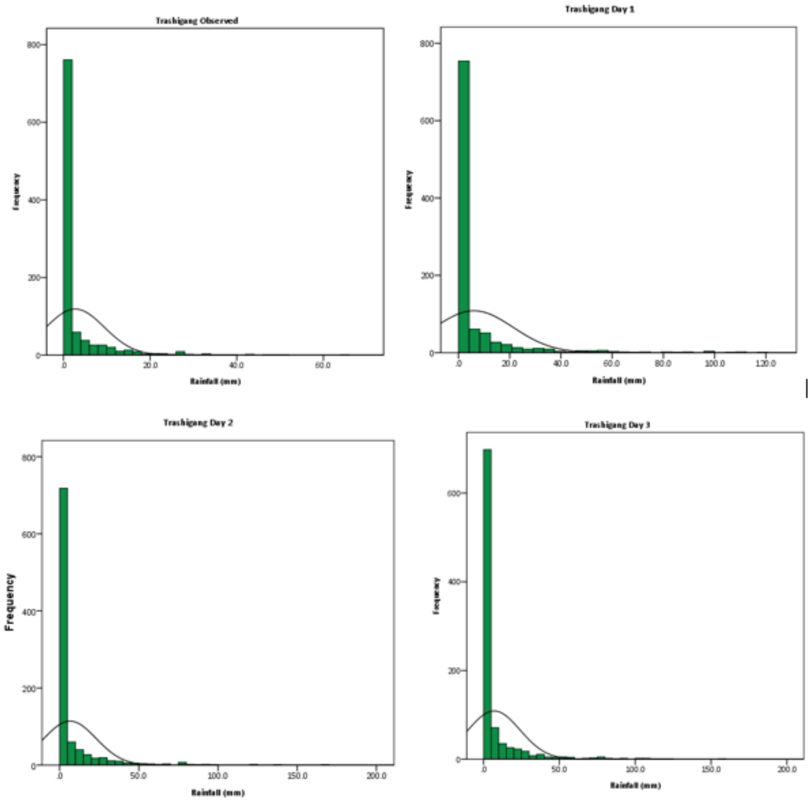

Figure 10: Histogram for Trashigang

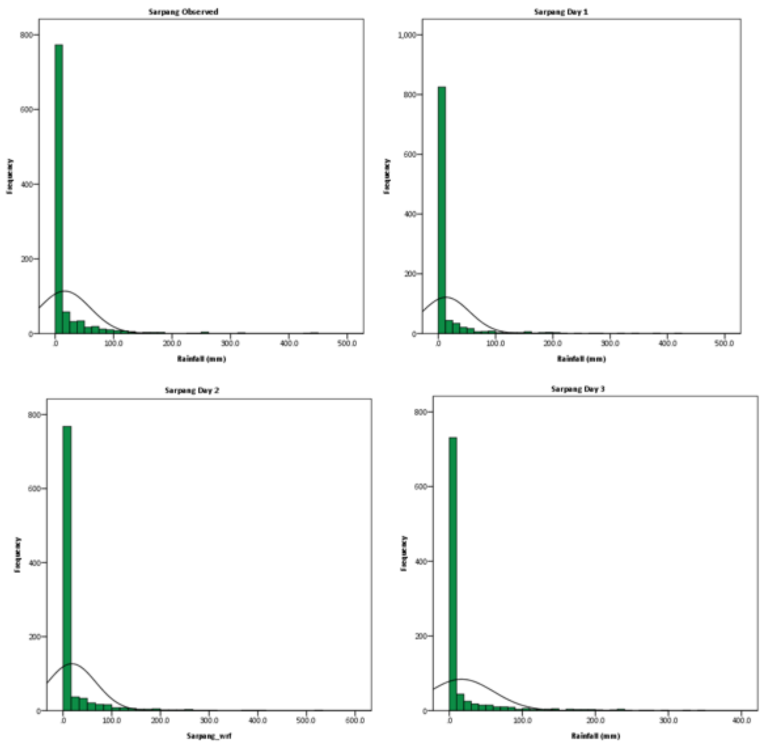

Figure 11: Histogram for Sarpang

The curve line tail tends to extend out to the right which illustrates to be rightly skewed for the Country. It demonstrates platykurtic kurtosis with the mean rainfall of 1.5 mm for observation (figure 9), 1.6 mm for Day 1, 1.7 mm for Day 2 and 2.0 mm for Day 3 for Thimphu. The extreme rainfall recorded is 44.0 mm while model extreme is 59.1 mm for Thimphu. The highest frequency of rainfall recorded is between 0-10 mm for model and observation.

The figure 10 shows the mean rainfall of 2.7 mm for observation, 6.0 mm for Day 1, 6.5 mm for Day 2 and 7.0 mm for Day 3 for Trashigang. The extreme rainfall recorded is 64.0 mm while model extreme is 167.1 mm for forecast Day 2. Therefore, model is over estimating the extreme rain for the Trashigang. The highest frequency of rainfall recorded is between 0-20 mm.

The figure 11 shows the mean rainfall of 14.4 mm for observation, 13.2 mm for Day 1, 18.0 mm for Day 2 and 16.6 mm for Day 3 for Sarpang. The extreme rainfall recorded is 432.0 mm while the model predicts 525.0 mm. The highest frequency of rainfall recorded is between 0-40 mm.

3.3 Monthly average temperature and rainfall for model and observation

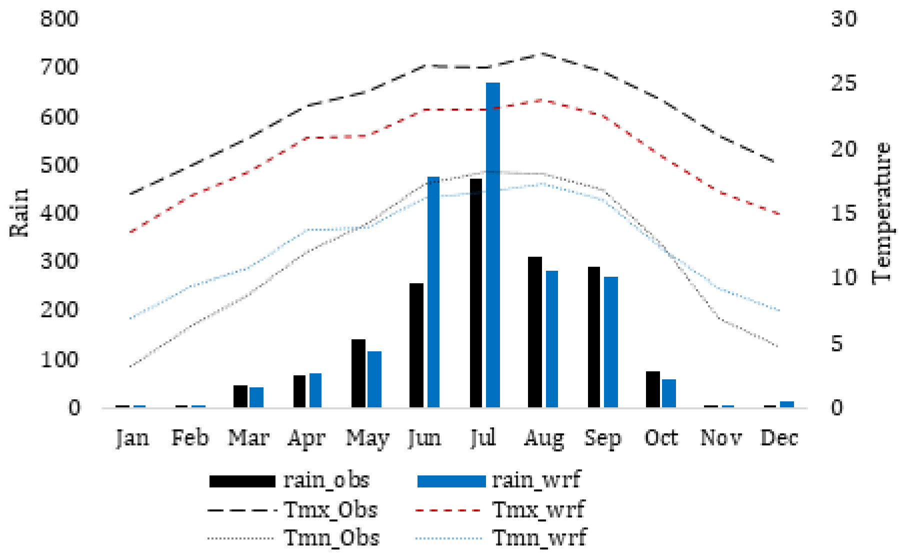

Figure 12: Monthly average for model and observation

The five focal point average (2016-2018) temperature and rainfall is plotted to see whether the model captures monthly variation of temperature and rainfall (figure 12). Overall, the model is able to capture the monthly variation for rainfall with peak rainfall during July however, there is bias of 200 mm. The rainfall bias is more for the month of June with 219 mm (table 7). The spring, autumn and winter rainfall is predicted well with minimal bias which is expected since these seasons has minimum rainfall total.

The model is able to follow the seasonal variation of the maximum and minimum temperature with persistent cooler bias for maximum, whereas, the model overestimates for the colder months and underestimates for the summer season.

Table 7: Monthly average for model and observation

Month	Temp Maximum			Temp Minimum			Rainfall		
	OBS	**WRF**	**Bias**	**OBS**	**WRF**	**Bias**	**OBS**	**WRF**	**Bias**
Jan	16.6	13.6	-3.0	3.3	6.9	3.7	5.1	4.0	-1.1
Feb	18.7	16.4	-2.3	6.3	9.4	3.1	6.0	2.8	-3.2
Mar	20.9	18.2	-2.7	8.9	10.7	1.9	46.0	41.3	-4.7
Apr	23.4	20.9	-2.5	12.0	13.7	1.7	67.1	73.5	6.3
May	24.4	21.1	-3.3	14.3	14.0	-0.3	140.4	116.4	-24.0
Jun	26.5	23.0	-3.5	17.3	16.3	-1.0	255.9	475.4	219.5
Jul	26.3	23.1	-3.1	18.2	16.7	-1.5	472.4	672.6	200.2
Aug	27.3	23.9	-3.5	18.0	17.3	-0.8	311.1	280.8	-30.3
Sep	26.0	22.6	-3.4	16.9	16.0	-0.9	290.2	271.2	-19.0
Oct	23.9	19.4	-4.5	12.7	12.4	-0.4	75.1	61.1	-14.0
Nov	21.1	16.7	-4.4	7.0	9.2	2.2	5.6	5.7	0.0
Dec	18.8	14.9	-3.9	4.8	7.5	2.8	5.9	13.9	8.0
Avg	**22.8**	**19.5**	**-3.3**	**11.6**	**12.5**	**0.9**	**140.1**	**168.2**	**28.1**

4 Conclusion

The three years analysis of the WRF output with the observation from 5 focal point of the country has been done for the 3 days forecast.

Overall, the model has good performance for the maximum temperature forecast with the bias of -3.1 °C for Bumthang, +0.6 °C for Tsirang, -1.3 °C for Sarpang, -5.3 °C for Trashigang and -1.9 °C for Thimphu. Lower RMSE is shown over the Tsirang and high over the Thimphu, however, there is increase of RMSE with the lead time throughout the stations. We have seen a very good correlation between the model and observations ranging from 0.7 to 0.9 for Bhutan. The stations located at the southern part of the country is performing better than the stations of west, east and central.

The model performed well for the minimum temperature forecast with the bias of -0.6 °C for Bumthang, +1 °C for Tsirang, -0.4 °C for Thimphu, -0.7 °C for Sarpang and +2.7 °C for Trashigang. There are noticeable problem in the winter season as model is failing to capture the extreme minimum temperature and when can see there are bias when looked into the plots and graphs. Lower RMSE is shown over the Sarpang and high over the Thimphu, however, RMSE is found to be more or less improving for all the stations. We could find very good correlation between model and observation ranging from 0.7 to 0.9 for Bhutan. The southern station is found to be performing well for minimum temperature than west, east and central Bhutan.

Rainfall events were captured well with the bias of ± 0.2 for all the stations, however, it was found to be increasing with the lead time. PAG, FAR, POD and F was found to be good for the stations located at the southern part of the country whereas, less performance was shown from the stations located at west and central part of the country.

The model is able to capture the monthly variation for rainfall with peak rainfall during July however, there is bias of 200 mm. The rainfall bias is more for the month of June with 219 mm. The spring, autumn and winter rainfall is predicted well with minimal bias. The model is able to follow the seasonal variation of the maximum and minimum temperature with persistent cooler bias for maximum, whereas, the model overestimates for the colder months and underestimates for the summer season.

5 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*

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 things 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 ration 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 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 is 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)]

Annexure-C

Contingency table

Day	Forecast Event	Observed Event Yes	Observed Event No	Marginal total
Day 1	Yes	176	118	294
Day 1	No	144	561	705
Day 1	Marginal total	320	705	
Day 2	Yes	146	118	258
Day 2	No	163	522	685
Day 2	Marginal total	309	634	
Day 3	Yes	137	114	218
Day 3	No	164	512	685
Day 3	Marginal total	301	635	

Table 8: Bumthang rainfall event for forecast days (Day 1, Day 2, Day 3)

Day	Forecast Event	Observed Event Yes	Observed Event No	Marginal total
Day 1	Yes	177	138	315
Day 1	No	103	581	684
Day 1	Marginal total	280	719	
Day 2	Yes	173	156	329
Day 2	No	86	528	614
Day 2	Marginal total	259	684	
Day 3	Yes	173	139	312
Day 3	No	93	527	620
Day 3	Marginal total	266	666	

Table 9: Tsirang rainfall event for forecast days (Day 1, Day 2, Day 3)

Day	Forecast Event	Observed Event Yes	Observed Event No	Marginal total
Day 1	Yes	111	134	245
Day 1	No	122	632	754
Day 1	Marginal total	233	766	
Day 2	Yes	113	134	247
Day 2	No	110	586	696
Day 2	Marginal total	223	720	
Day 3	Yes	122	134	256
Day 3	No	98	578	676
Day 3	Marginal total	220	712	

Table 10: Thimphu rainfall event for forecast days (Day 1, Day 2, Day 3)

Day	Forecast Event	Observed Event Yes	Observed Event No	Marginal total
Day 1	Yes	177	157	334
Day 1	No	120	545	665
Day 1	Marginal total	297	702	
Day 2	Yes	183	161	344
Day 2	No	105	494	599
Day 2	Marginal total	288	655	
Day 3	Yes	191	148	339
Day 3	No	94	499	593
Day 3	Marginal total	285	647	

Table 11: Trashigang rainfall event for forecast days (Day 1, Day 2, Day 3)

Day	Forecast Event	Observed Event Yes	Observed Event No	Marginal total
Day 1	Yes	237	86	323
Day 1	No	256	520	676
Day 1	Marginal total	393	606	
Day 2	Yes	245	88	333
Day 2	No	132	477	610
Day 2	Marginal total	377	565	
Day 3	Yes	235	81	316
Day 3	No	137	479	616
Day 3	Marginal total	372	560	

Table 12: Sarpang rainfall event for forecast days (Day 1, Day 2, Day 3)