24 HOUR WEATHER FORECAST VERIFICATION

FOR YEAR 2018 AND 2019

Weather and Climate Services Division

National Center for Hydrology and Meteorology

2020

Table of Contents

Table of Figures and Tables

1. Introduction

Weather forecast services in Bhutan started in 2007. Currently, the National Center for Hydrology and Meteorology (NCHM) issues the daily weather forecast for the next 24 hours to the nation through the national television Bhutan Broadcasting Service (BBS) and print media Kuensel. The forecast comprises the expected surface maximum and minimum temperature in degrees Celsius (°C) and a general outlook for the next 24 hours.

With advancements in weather and climate knowledge and information, there is an increasing demand for these services. NCHM is mandated to provide accurate, reliable, timely and consistent weather information to the nation. This report will;

a. Validate, using simple statistical methods, the accuracy of the daily weather forecast for the next 24 hours for variables of surface maximum and minimum temperature in degrees Celsius (°C) and the event of rainfall for the year 2018 and 2019 (1 January 2018 to 31 December 2019), by comparing the forecast data with the observation data from the 20 Agrometeorological stations (Class A)
b. Provide a guidance for weather forecasting for variables of surface maximum and minimum temperature and the event of rainfall

2. Data and Methodology

2.1 Observation data

Meteorological variable of surface temperature and event of rainfall is used for the verification of the daily weather forecast for the next 24 hours. There are 20 Agrometeorological stations (Class A) that represent the country, which are identified as the focal point of weather forecasting for Bhutan. These stations are manned by regular NCHM staff and report data to the National Weather and Flood Warning Center (NWFWC) 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 oktas are reported. Besides these variables Class A stations also measure other weather variables.

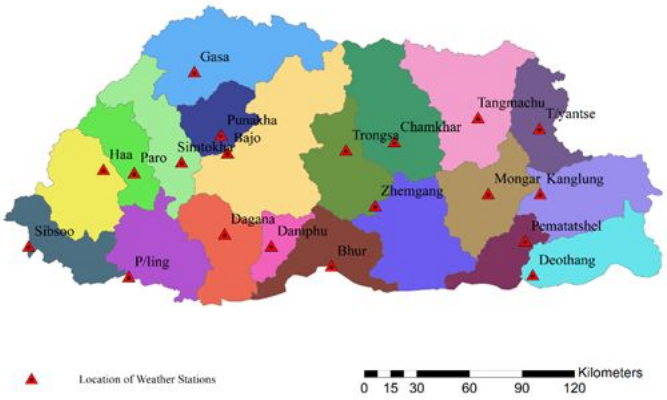

Figure 2.1: Location of 20 Agromet Stations (Class A)

Observed surface maximum and minimum temperature in degrees Celsius (°C) and the event of rainfall data from the 20 Agrometeorological stations is compared with the daily weather forecast for the next 24 hours data of the year 2018 and 2019 (1 January 2018 to 31 December 2019) for this verification analysis. Verification is performed for individual station points.

However, since the observation is recorded from 9:00 AM till the next 24 hours and the forecast is predicted from 4:00 PM till the next 24 hours, there are limitations to the analysis report.

2.2. 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 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 statistical analysis are 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)

2.3. Dichotomous variable analysis

We defined the event before creating a dichotomous variable and accordingly, the contingency table for rainfall is prepared (refer Table 2.1) with daily weather forecast for the next 24 hours as 'Event Forecast (yes/no)' and observed rain from the stations as 'Event Observed (yes/no)' to collect a match set of forecast and observation. Rainy days are termed when the observed and forecast records 1 mm or more rainfall in a day.

	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 2.1: Contingency table for dichotomous variable analysis

- '**Hit**' is defined by the 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 when actually there was a 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.

2.3.1 Calculating scores using the contingency table

From the contingency table generated from forecast 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

Verification of the daily 24 hours weather forecast is performed for 20 individual station points, where the observation from these station points are compared with the individual forecast.

3.1. Continuous Variable

The analysis of the continuous variable is represented in time series, scatter plots and box plots for each of the 20 stations. Mean, Median, Standard Deviation (SD), Mean Error (ME), and Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and correlation has been calculated.

3.1.1. Maximum Temperature Analysis

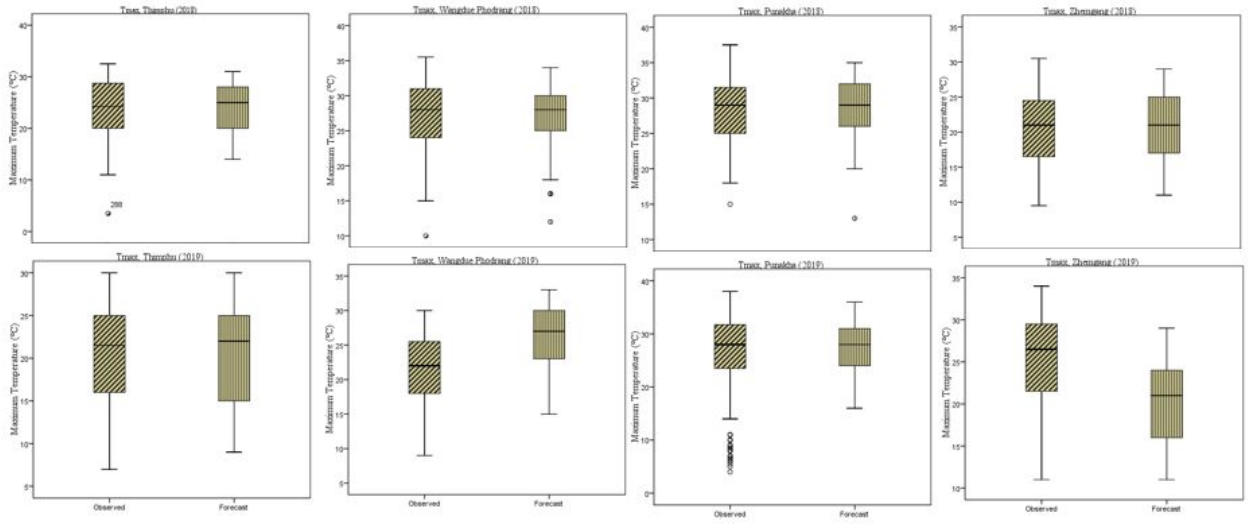

Figure 3.1: Boxplot of maximum temperature for selected 4 station points

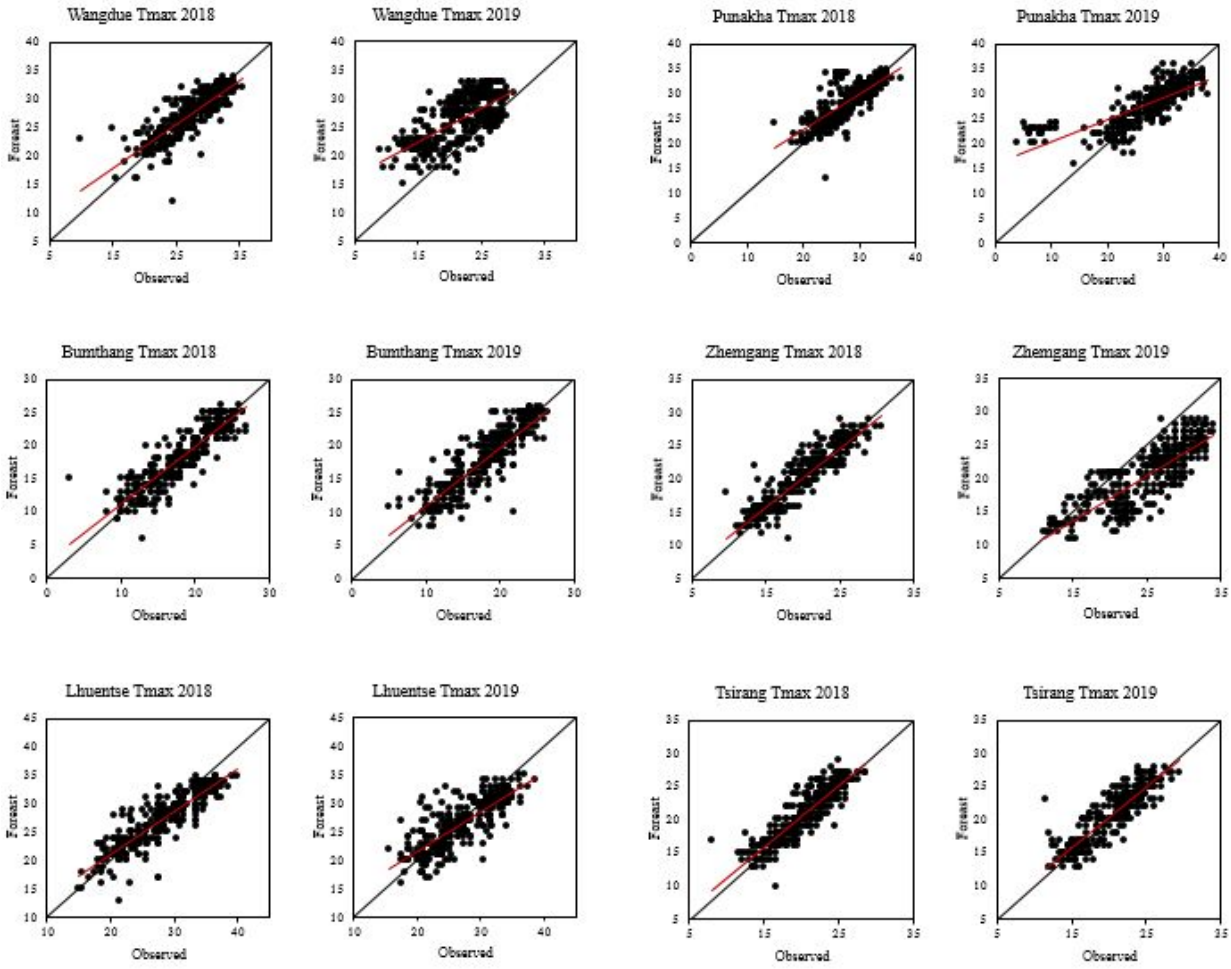

Figure 3.2: Scatter plot of maximum temperature for selected 6 station points

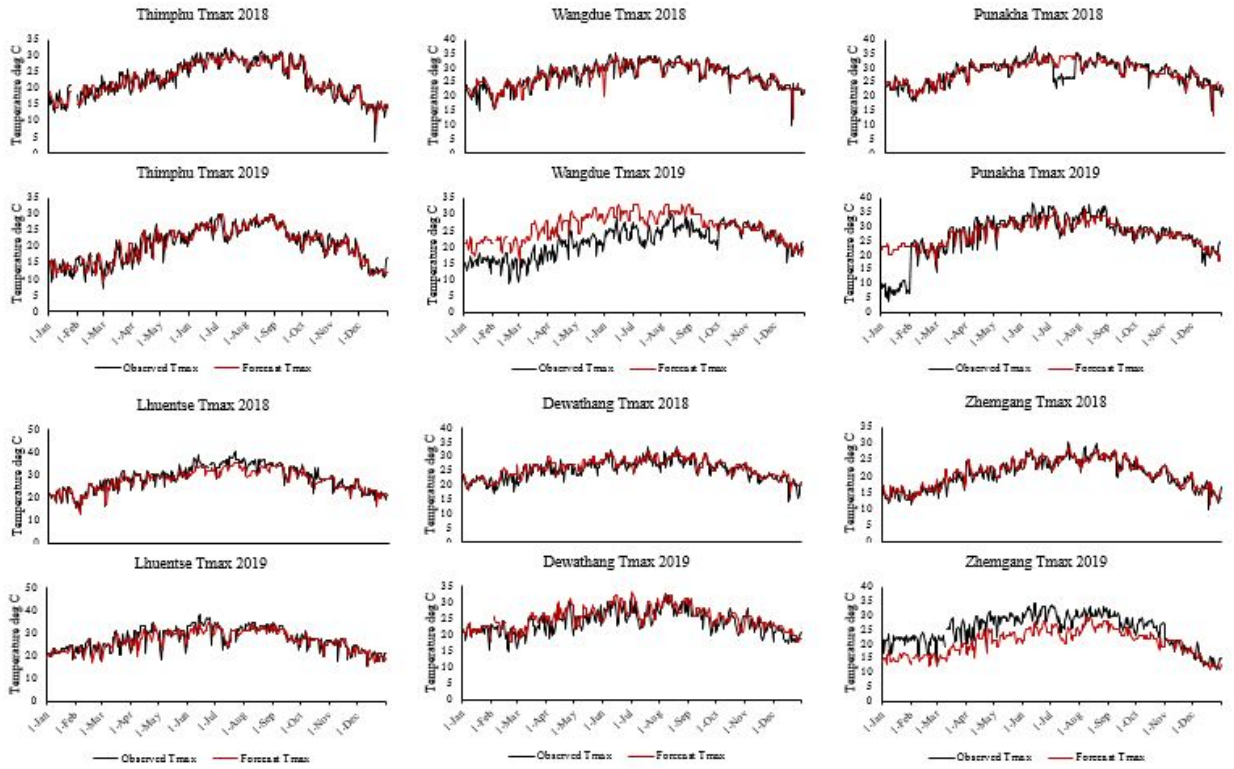

Figure 3.3: Time series plot of maximum temperature for selected 6 stations points

The box plot analysis for maximum temperature for both the years 2018 and 2019 show similar dispersion of values of observation to forecast indicating the forecast was captured for most of the stations except for Wangdue where the forecast dispersion is lesser than the observation in 2018. The observation dispersion is lesser compared to the forecast at Tashi Yangtse during the same year. The maximum temperature was under predicted for Zhemgang in 2019. It was over predicted for Wangdue, Gelephu and Mongar for both the years. Largest outliers are indicated in Punakha in 2019.

The scatter plot analysis for both years 2018 and 2019 shows positive linear correlation association between the maximum temperatures forecast and observed for all the station points. 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 was seen at most of the station points. However, larger dispersions can be seen at Wangdue, Punakha, Lhuentse and Dagana for both the years. The maximum temperature was over predicted for Wangdue and Punakha in 2019, nevertheless the maximum temperature was under predicted for Zhemgang in 2019. Punakha indicated the largest uncaptured extreme values.

The time series analysis shows that the maximum temperature forecast was relatively well for most of the stations for both the years 2018 and 2019. The maximum temperature was over predicted for Deothang for both the year with an average of +1 °C and for Haa for both the year during summer (June and July). The largest maximum temperature forecast was for Wangdue in

2019 with an average of +5 °C from January to October. The maximum temperature was under predicted for Punakha from July- August 2018 and January to February in 2019. It was also under predicted for Lhuentse during the summer season for both the years. The highest under prediction of maximum temperature was in Zhemgang from January to November in 2019 with an average of -5 °C.

Year		Mean	Medium	SD	ME	MAE	RMSE	Correlation
				Thimphu				
2018	Forecast	28.8	22.0	5.2	-0.2	1.8	2.2	0.9
	Observation	22.9	22.3	5.5				
2019	Forecast	20.5	22.0	5.4	0.0	1.7	2.3	0.9
	Observation	20.5	21.5	5.5				
				Wangdue				
2018	Forecast	27.3	28.0	3.9	-0.1	1.6	2.3	0.8
	Observation	27.3	27.5	4.3				
2019	Forecast	26.2	27.0	4.1	4.7	5.0	5.9	0.7
	Observation	21.5	22.0	4.6				
				Bumthang				
2018	Forecast	18.1	18.0	4.3	0.0	1.5	2.1	0.9
	Observation	18.1	18.0	4.5				
2019	Forecast	18.1	18.0	4.5	0.1	1.7	2.3	0.9
	Observation	18.1	18.0	4.5				
				Zhemgang				
2018	Forecast	20.8	21.0	4.6	0.3	1.3	1.8	0.9
	Observation	20.5	20.5	4.5				
2019	Forecast	20.5	21.0	4.5	-4.6	4.8	5.5.	0.8
	Observation	25.1	6.3	5.5				
				Lhuentse				
2018	Forecast	27.7	28.0	4.7	-1.1	2.1	2.7	0.9
	Observation	28.6	28.5	5.5				
2019	Forecast	26.8	27.0	4.3	-0.6	2.4	3.4	0.8
	Observation	27.5	27.5	4.9				

Table 3.1: Analysis table for maximum temperature for selected 5 station points

For maximum temperature in the year 2018, the bias was negligible for Bumthang and Phuentsholing. It was under predicted for Thimphu, Wangdue, Tashigang and Lhuentse with bias ranging from -1.1 to -0.1. Rest of the station points were over predicted with bias ranging up to 0.7, which was recorded from Deothang. In the year 2019, the bias was negligible for Thimphu, Dagana and Samtse. It was under predicted for Zhemgang, Mongar, Lhuentse, Phuentsholing and Tashi Yangtse with bias ranging from -4.6 for Zhemgang to -0.1. Rest of the station points were over predicted with bias ranging up to 4.7 for Wangdue.

Paro, Gasa and Dagana showed decrease in RMSE with an average error value of 0.1. Rest of the station points indicated an increase in RMSE with error values ranging up to 4 for Wangdue and Zhemgang.

3.1.2 Minimum Temperature Analysis

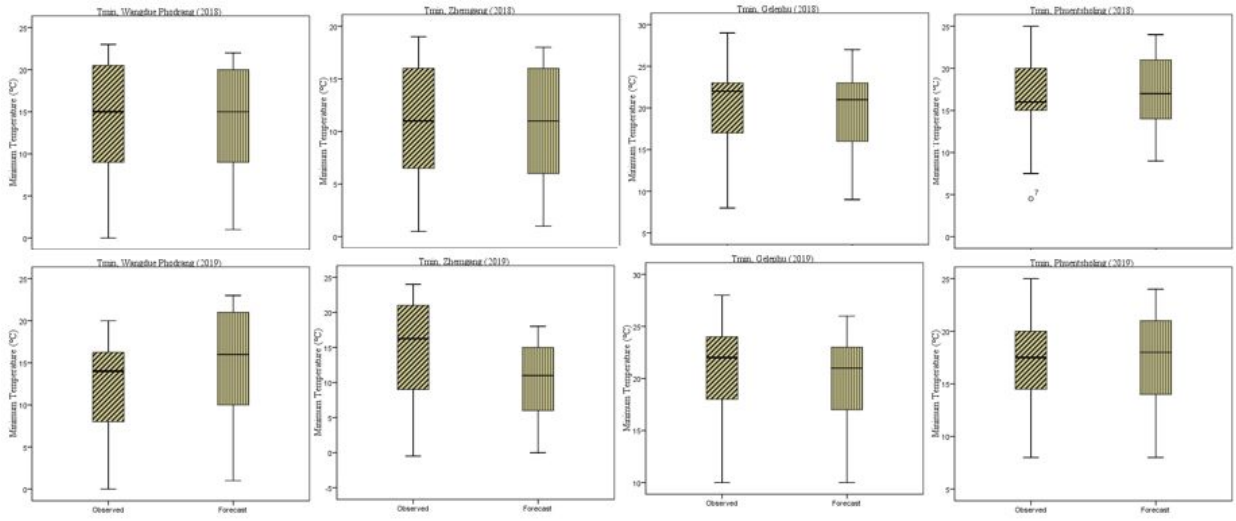

Figure 3.4: Box plot of minimum temperature for selected 4 station points

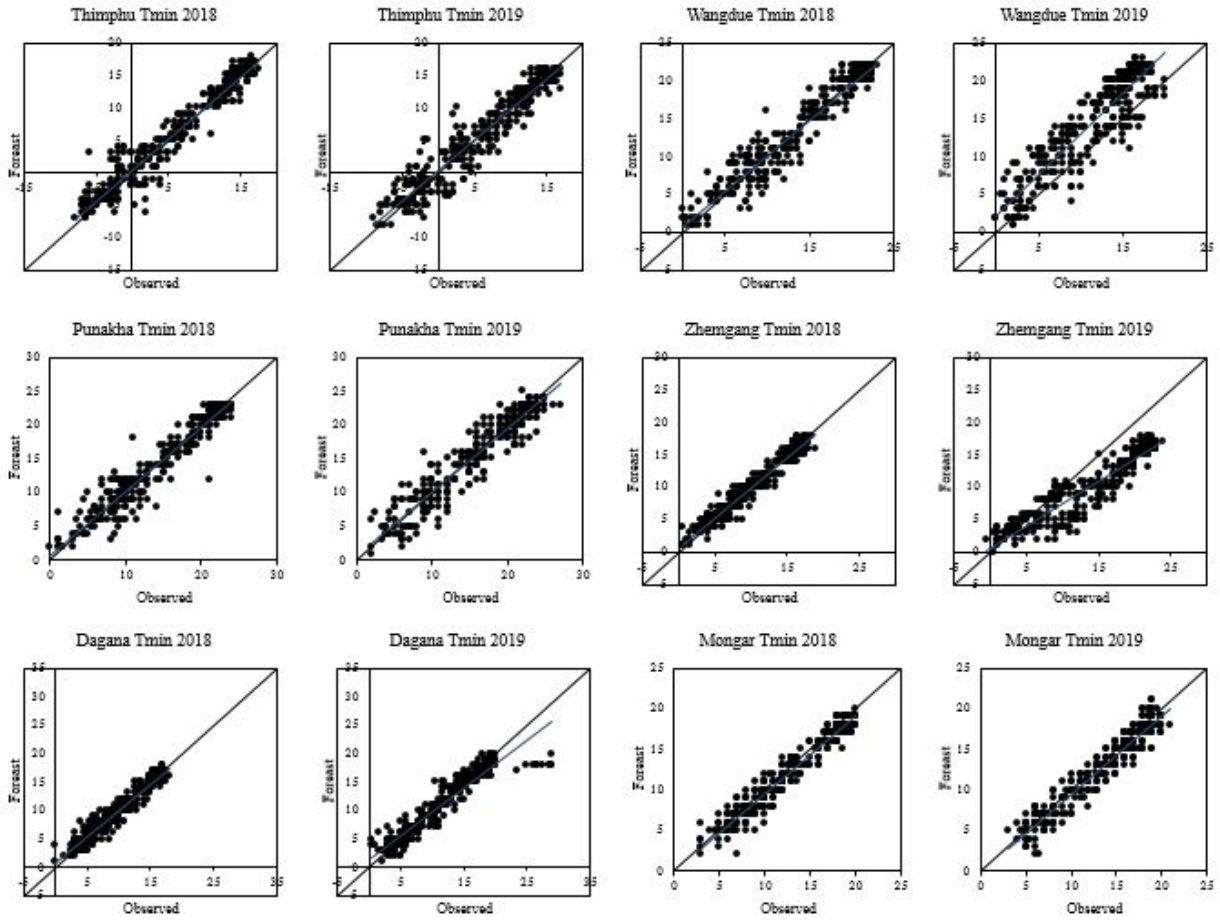

Figure 3.5: Scatter plot of minimum temperature for selected 6 stations points

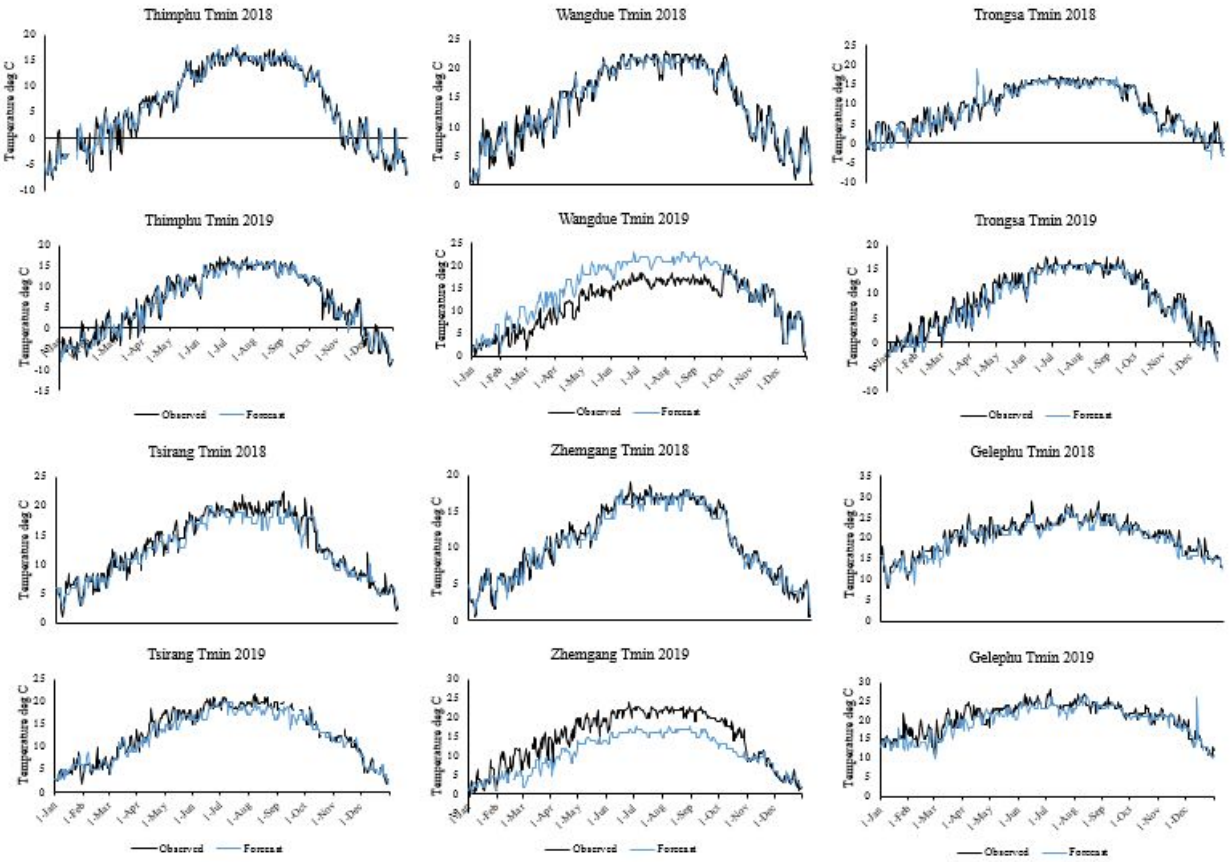

Figure 3.6: Time series plot of minimum temperature for selected 6 stations points

The box plot analysis for minimum temperature for both the years 2018 and 2019 shows similar dispersion of values of observation to forecast indicating the forecast was captured for most of the stations except for Wangdue and Phuentsholing where the forecast dispersion is larger than the observation in 2019 and for both the years respectively. The forecast dispersion is lesser to the observation for Zhemgang in 2019. The minimum temperature was under predicted for Zhemgang, Gelephu, Samtse and Pema Gatshel in 2019. However it was over predicted for Wangdue in 2019.

The scatter plot analysis for both the years 2018 and 2019 shows positive linear correlation association between the minimum temperatures forecast and observed for most of the stations. There is both positive and negative linear correlation in Thimphu, Paro, Haa, Gasa, Bumthang, Trongsa and Tashi Yangtse. 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 was seen at most of the station points. However, larger dispersions can be seen at Wangdue and Punakha for both the years. The minimum temperature was over predicted for

Wangdue in 2019. It was under predicted for Dagana and Zhemgang in 2019. Dagana indicated the largest uncaptured extreme values in 2019.

The time series shows that the minimum temperature forecast was relatively well for most of the stations for both the years 2018 and 2019. The minimum temperature was over predicted for largely at Wangdue from February to October in 2019 with an average of +5°C. It was under predicted by a few values of -1 to 3 °C at Trongsa during May and July in 2019, Tsirang during summer season in both the year, and Gelephu during summer season in 2019. The highest under prediction was in Zhemgang from February to November in 2019 with an average of -5 °C.

Year		Mean	Medium	SD	ME	MAE	RMSE	Correlation
				Thimphu				
2018	Forecast	6.4	7.0	7.5	0.1	1.3	1.9	1.0
	Observation	6.3	6.5	4.6				
2019	Forecast	6.3	7.0	7.2	-0.1	1.5	2.0	1.0
	Observation	6.5	7.5	7.4				
				Phuentsholing				
2018	Forecast	17.8	19.0	4.1	-0.2	0.9	1.2	1.0
	Observation	18.0	18.5	4.1				
2019	Forecast	17.9	18.0	4.0	-0.1	1.0	1.3	0.9
	Observation	18.0	19.0	4.1				
				Zhemgang				
2018	Forecast	10.6	11.0	5.0	-3.9	4.2	4.8	0.9
	Observation	14.4	16.0	7.0				
2019	Forecast	10.7	11.0	4.8	-0.2	0.9	1.1	1.0
	Observation	10.8	11.0	5.1				
				Gelephu				
2018	Forecast	19.6	21.0	4.1	-1.0	1.6	2.3	0.9
	Observation	20.7	22.0	3.9				
2019	Forecast	19.8	21.0	4.0	-0..7	1.3	1.7	0.9
	Observation	20.5	22.0	4.1				
				Deothang				
2018	Forecast	16.6	17.0	4.0	-0.2	1.0	1.4	09
	Observation	16.7	17.5	4.1				
2019	Forecast	16.5	17.0	4.1	0.1	1.0	1.3	0.9
	Observation	16.3	16.5	4.2				

Table 3.2: Analysis table for minimum temperature for selected 5 station points

For the year 2018, the minimum temperature was under predicted for most of the station points with the highest value of -3.9 at Zhemgang. It was over predicted for Wangdue, Tashigang and Phuentsholing with a bias ranging from 0.1 to 3.0. For the year 2019, the minimum temperature was negligible for Haa. It was over predicted for a few stations such as Thimphu, Paro, Lhuentse, Deothang and Phuentsholing with highest bias of 0.3 in Phuentsholing. Rest of the station points were under predicted with the largest bias of -0.7 in Gelephu.

The RMSE for Trongsa, Gelephu, Tashi Yangtse and Samtse shows an increase with an average error value of 0.1. Rest of the station points represent a decrease in the RMSE with error values ranging up to 4 in Zhemgang. Haa and Gasa show no change in RMSE.

3.2 Dichotomous Variables Analysis

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

Year	B	POD	FAR	PAG	F	KSS	HSS	TS
				Gasa				
2018	1.3	0.9	0.3	0.7	0.5	0.4	0.3	0.7
2019	1.3	0.9	0.3	0.7	0.4	0.5	0.3	0.7
				Thimphu				
2018	2.9	0.9	0.7	0.3	0.5	0.4	0.4	0.3
2019	2.8	0.9	0.7	0.3	0.5	0.4	0.4	0.3
				Bumthang				
2018	2.2	0.9	0.6	0.4	0.5	0.4	0.3	0.4
2019	2.3	0.9	0.	0.4	0.5	0.4	0.4	0.4
				Tashigang				
2018	2.2	1.0	0.6	0.4	0.5	0.5	0.4	0.4
2019	2.1	1.0	0.5	0.5	0.5	0.5	0.4	0.4
				Sarpang				
2018	1.7	0.9	0.4	0.6	0.4	0.5	0.4	0.5
2019	1.4	0.9	0.3	0.7	0.4	0.6	0.4	0.6

Table 3.3: Computed scored for rainfall using contingency table for selected 5 stations

The Frequency Bias Index (B) ranges up to value 2.9 for Thimphu and Paro in 2018 and 2.8 for Thimphu in 2019. Most of the station points indicate a decrease in the bias index from 2018 to 2019, with highest value 0.8 for Wangdue. However, bias indices for Gasa, Bumthang and Samtse remained the same over the year.

The Probability of Detection (POD), sometimes called Hit rate remained the same for most of the stations over the year between the value 0.9-1.0 meaning the forecast was able to capture the event of rainfall 90 and 100 % respectively. There was an increase in the POD at Gasa, Trongsa, Mongar, Gelephu, Dewathang and Phuentsholing from 0.9-1.0. However, Wangdue and Punakha showed decrease in POD from 0.9-0.8 and Tashi Yangtse with value 1.0-0.9.

There is decrease in False Alarm Ratio (FAR) and Post agreement (PAG) for Gasa, Trongsa, Wangdue, Dagana, Zhemgang, Tashigang, Lhuentse, Gelephu, Dewathang, Pema Gatshel and Phuentsholing with 0.1 value over the year, indicating 10% decrease in the forecast of rainfall event captured. However the FAR and PAG for rest of the station points remained the same for values between 0.3- 0.7. Lowest FAR value of 0.2 with the highest PAG value of 0.8 was indicated at Gasa in 2019. The highest FAR with the lowest PAG was shown at Thimphu and Paro for both the years and for Wangdue in 2018 with a value of 0.7 and 0.3 respectively.

The False alarm rate (F) remained the same for Trongsa and Dewathang with rate 0.5. Rest of the station points showed an increase with a rate of 0.1 indicating 10% increase in forecast unable to capture the event of rainfall. The rate ranged from 0.3 at Samtse in 2018 to 0.7 in Bumthang in 2019.

There is a decrease in the Hanssen & Kuiper`s skill (KSS) and Heidke skill score (HSS) for most of the stations with score 0.1, indicating a 10 % decrease in the forecast was able to capture the rainfall event for most of the station over the year. The lowest score 0.2 was recorded at Paro, Gasa, Bumthang, Tsirang and Lhuentse in 2019. The highest score of 0.6 was at Samtse in 2018.

Thimphu, Paro, Haa, Gasa, Bumthang, Tsirang, Punakha, Dagana and Samtse showed the same Threat score (TS) over the years with score range from 0.3, indicating 30 % forecast was able to capture the rainfall event with lowest score of 0.3 at Thimphu and Gasa and highest 0.7 for Samtse. Rest of the station points show an increase in the score by 0.1.

4. Conclusion

Validation of the daily weather forecast for the next 24 hours for variables of surface maximum and minimum temperature in degrees Celsius (°C) and the event of rainfall for the year 2018 and 2019 (1 January 2018 to 31 December 2019) by comparing the forecast data with the observation data from the field using simple statistical methods has been done.

The forecast for surface maximum temperature was captured for most of the stations for both the years 2018 and 2019. However, it was over predicted for Wangdue, Gelephu, Deothang, Haa and Mongar for both the years and Punakha in 2019. The highest maximum temperature over prediction was at Wangdue in 2019 with an average of +5 °C from January to October. The maximum temperature was under predicted for Punakha and Tashigang in 2018, Lhuentse for both the years and Tashi Yangtse in 2019. The largest maximum temperature under prediction was at Zhemgang from January to November in 2019 with an average of -5 °C.

Paro, Gasa and Dagana showed decrease in RMSE with an average error value of 0.1 over the year but the rest of the station points indicated increase in RMSE with error value ranging up to 4 for Wangdue and Zhemgang. Punakha indicated the largest uncaptured extreme maximum temperature in 2019.

The forecast for surface minimum temperature for both the years 2018 and 2019 was captured for most of the stations. The minimum temperature was over predicted for Wangdue, Tashigang and Phuentsholing in 2019. The minimum temperature was over predicted for largely at Wangdue from February to October in 2019 with an average of +5°C. However, it was under predicted for Zhemgang, Dagana, Gelephu, Samtse and Pema Gatshel in 2019. The highest under prediction was in Zhemgang from February to November in 2019 with an average of -5 °C.

The RMSE for Trongsa, Gelephu, Tashi Yangtse and Samtse shows an increase with an average error value of 0.1. Rest of the station points represent a decrease in the RMSE with error values ranging up to 4 in Zhemgang. Haa and Gasa show no change in RMSE. Dagana indicated the largest uncaptured extreme minimum temperature in 2019.

The forecast for rainfall events were captured for both the years 2018 and 2019 with Probability of Detection (POD) value of 0.9-1.0 respectively. Most of the station points indicate decrease in the bias index from 2018 to 2019 with skill score 0.1 indicating 10 % decrease in the forecast unable to capture the rainfall event over the year. The lowest skill score 0.2 was recorded at Paro, Gasa, Bumthang, Tsirang and Lhuentse in 2019 and the highest 0.7 for Samtse in both the years.

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*
- WRF Model Verification, 2019, *National Center for Hydrology and Meteorology, Thimphu, Bhutan*

6. Annexure- 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 on the manicure of errors. The value ranges from -∞ to +∞. The perfect score is equal to 0.

$ME = (1/N) \sum (fi-fo)$

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.

$MAE = (1/N) \sum |fi-fo|$

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.
$RMSE = (1/N) \sum (fi-fo)^2$

2. Dichotomous variables

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)

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)

The Threat Score (TS) 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).

$KSS=TSS=POD-F$

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