# 24 HOUR WEATHER FORECAST VERIFICATION
# FOR YEAR 2021

**Weather and Climate Services Division**

**National Center for Hydrology and Meteorology**

**2021**

## 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 from 1st January 2016 to 31st December 2020, 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.

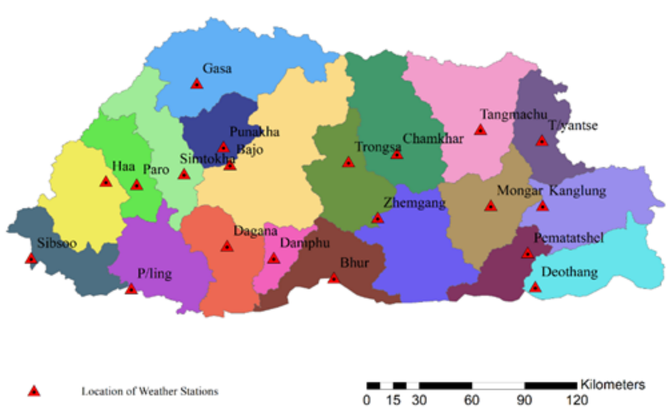


Figure 2.1: Location of 20 Agromet Stations (Class A)

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 from 1st January 2016 to 31st December 2020 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 lines 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 observations 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

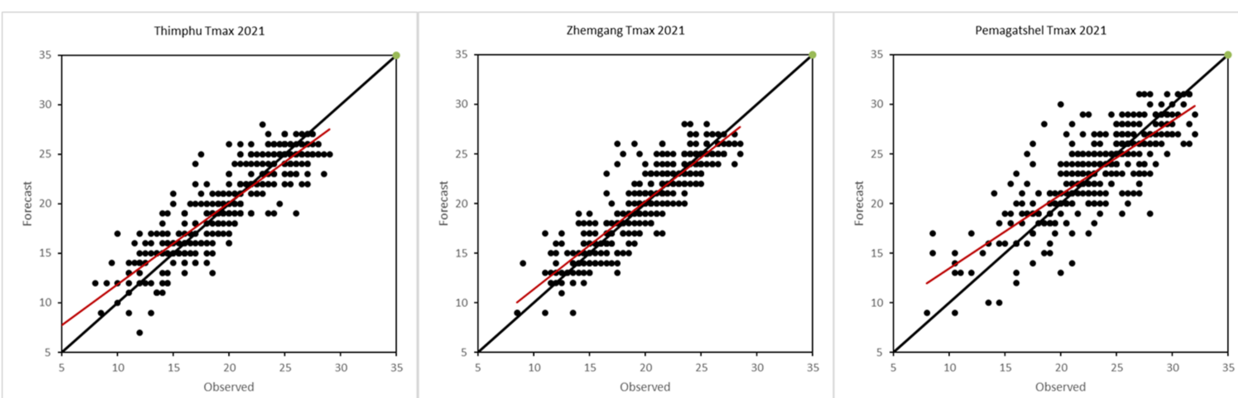


Figure 3.2: Scatter plot of maximum temperature for selected 3 station points

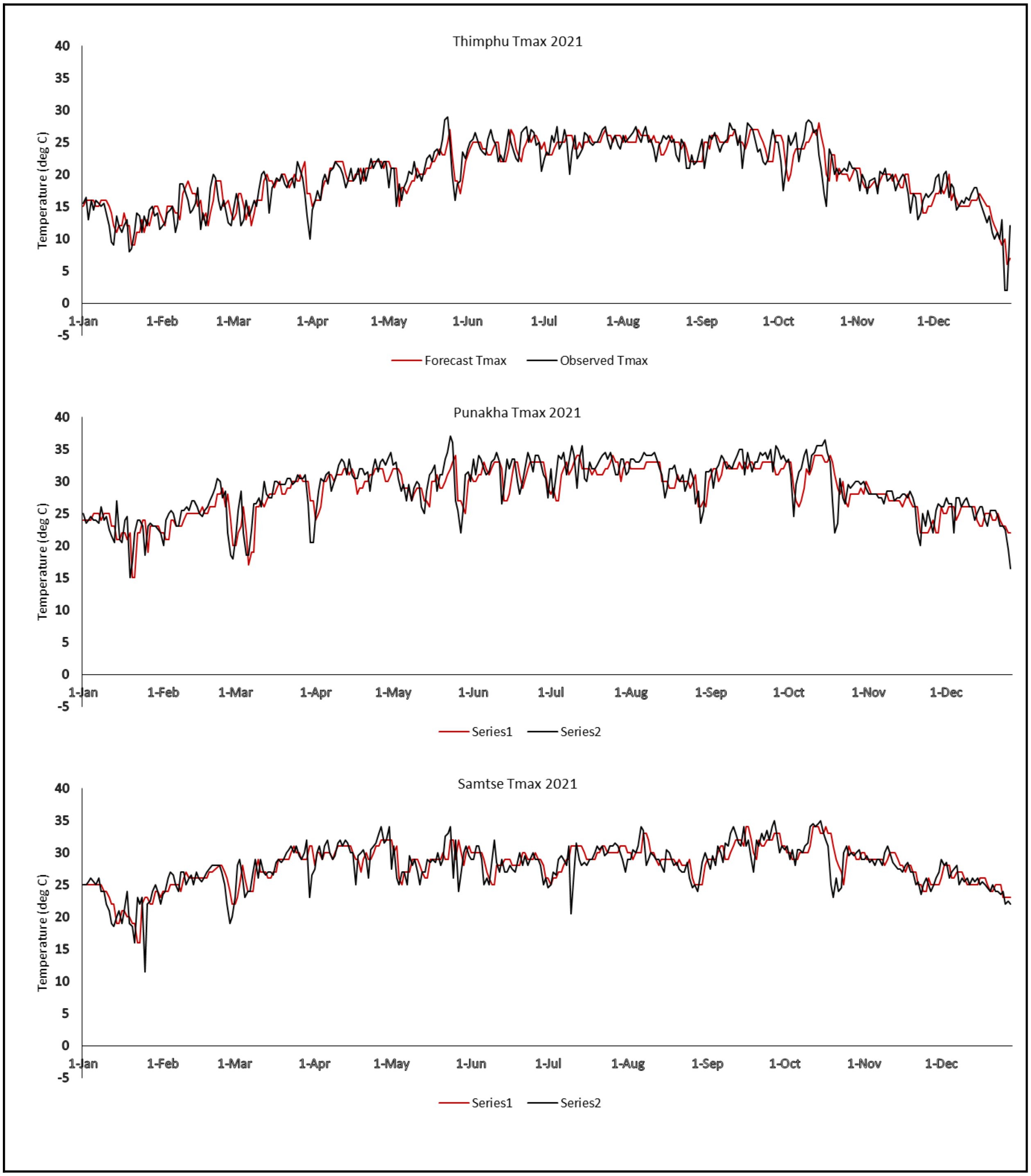


Figure 3.3: Time series plot of maximum temperature for selected 3 stations points

The scatter plot analysis for 2021 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. Larger dispersions can be seen at Paro, Mongar, Pema Gatshel, and Trashiyangtse for all the years. The prediction was relatively well for Zhemgang.

The time series analysis shows that the maximum temperature forecast was relatively well for most of the stations for 2021. The time series graphs shows that for most of the stations extreme maximum temperature was not well captured by the forecast. The forecast was able to well capture the seasonal changes. The maximum temperature was under predicted most of the time for Phuentsholing and Punakha.

Table 3.1: Analysis table for maximum temperature for selected 2 station points

| Year | | Mean | Median | SD | ME | MAE | RMSE | Correlation |
|---|---|---|---|---|---|---|---|---|
| | | | | Thimphu | | | | |
| 2021 | Forecast | 20.02 | 21.0 | 4.6 | 0.1 | 1.8 | 2.3 | 0.87 |
| | Observation | 20.02 | 20.5 | 5.0 | | | | |
| | | | | Bumthang | | | | |
| 2021 | Forecast | 18.2 | 18.0 | 4.2 | -0.2 | 1.8 | 2.4 | 0.84 |
| | Observation | 18.3 | 18.5 | 4.3 | | | | |

For maximum temperature in the year 2021, the bias was negligible for Sarpang and Dagana. Punakha had the highest under prediction, with a bias of -0.6 and Haa had the highest over prediction, with a bias of 0.6.

The RMSE for most of the stations ranges from 2 to  2.9. Paro has the highest RMSE error and Zhemgang has the lowest RMSE error.

### 3.1.2 Minimum Temperature Analysis

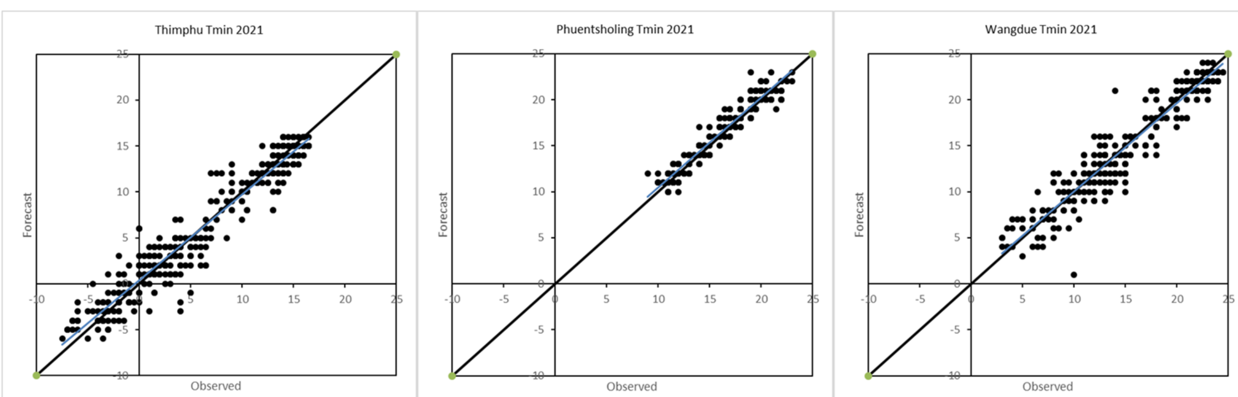


Figure 3.5: Scatter plot of minimum temperature for selected 3 stations points

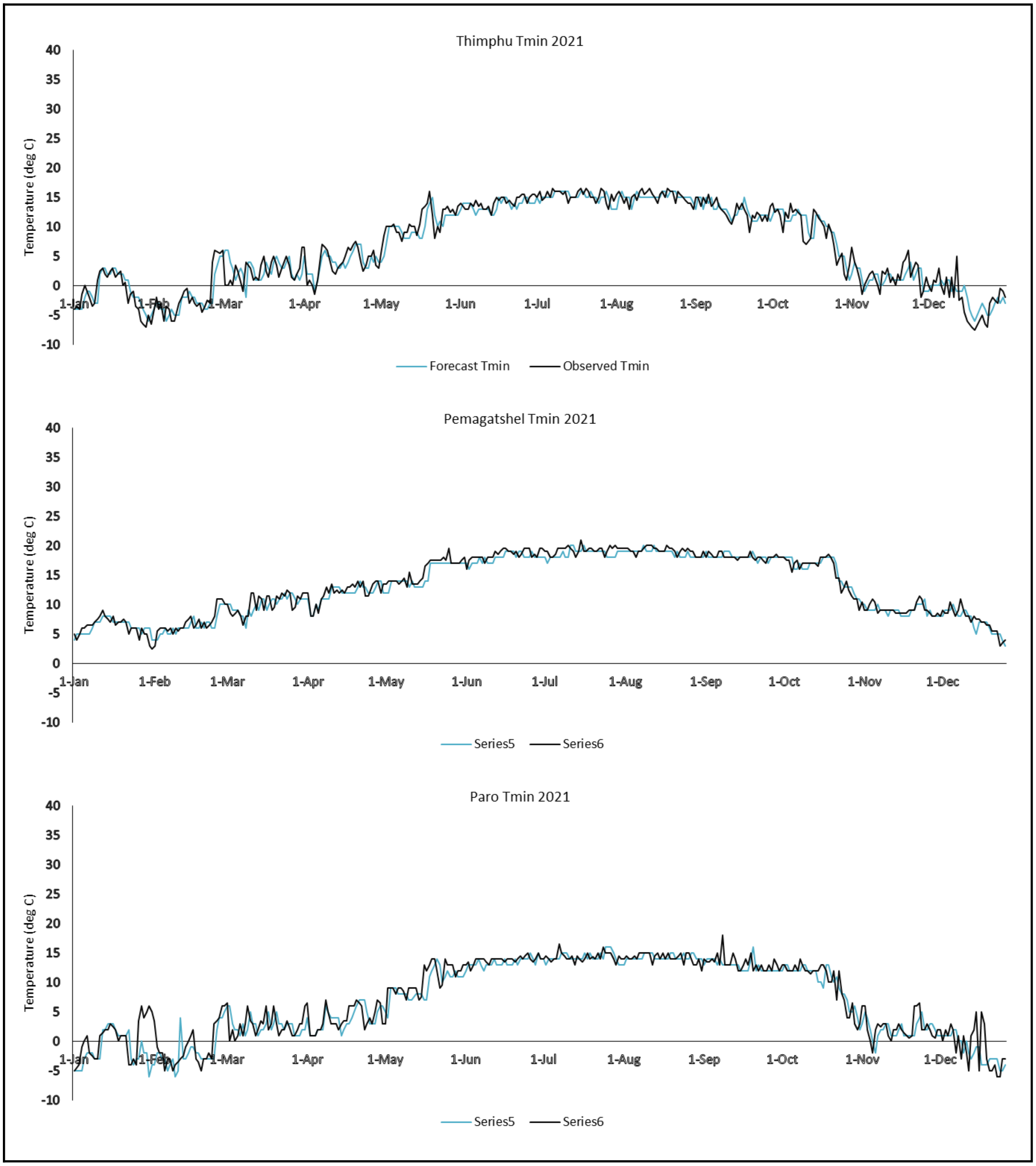

Figure 3.6: Time series plot of minimum temperature for selected 3 stations points

The scatter plot analysis for 2021 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. Larger dispersions can be seen at Bumthang, Haa and Paro. The prediction was relatively well for Gasa, Dagana, Zhemgang, Mongar, Samdrup Jonkhar, Pema gatshel and Phuentsholing

The time series analysis shows that the minimum temperature forecast was relatively well for most of the stations for 2021. The time series graphs shows that for most of the stations extreme maximum temperature was not well captured by the forecast. The forecast was able to well capture the seasonal changes. The minimum temperature was under-predicted most of the time for Paro and Gasa during the winter season.

Table 3.2: Analysis table for minimum temperature for selected 2 station points

| Year | | Mean | Median | SD | ME | MAE | RMSE | Correlation |
|---|---|---|---|---|---|---|---|---|
| | | | | Thimphu | | | | |
| 2021 | Forecast | 20.02 | 21.0 | 4.6 | 0.1 | 1.8 | 2.3 | 0.87 |
| | Observation | 20.02 | 20.5 | 5.0 | | | | |
| | | | | Bumthang | | | | |
| 2021 | Forecast | 18.2 | 18.0 | 4.2 | -0.2 | 1.8 | 2.4 | 0.84 |
| | Observation | 18.3 | 18.5 | 4.3 | | | | |

For maximum temperature in the year 2021, the bias was negligible for Samtse and Tashigang. Sarpang, Tsirang and Trongsa had the highest under prediction, with a bias of -0.6 and Phuentsholing had the highest over prediction, with a bias of 0.3.

The RMSE for most of the stations ranges from 1 to 2.3. Haa has the highest RMSE error and Mongar has the lowest RMSE error.

### 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.

Table 3.3: Computed scored for rainfall using contingency table

| **Stations** | **B** | **POD** | **FAR** | **PAG** | **F** | **KSS** | **HSS** | **TS** |
|---|---|---|---|---|---|---|---|---|
| **Bumthang** | 2.25 | 0.97 | 0.57 | 0.43 | 0.51 | 0.47 | 0.34 | 0.43 |
| **Dagana** | 2.24 | 0.99 | 0.56 | 0.44 | 0.45 | 0.54 | 0.39 | 0.44 |
| **Gasa** | 1.10 | 0.88 | 0.20 | 0.80 | 0.33 | 0.55 | 0.56 | 0.72 |
| **Haa** | 1.96 | 0.91 | 0.54 | 0.46 | 0.48 | 0.43 | 0.34 | 0.44 |
| **Lhuentse** | 2.47 | 0.90 | 0.64 | 0.36 | 0.51 | 0.38 | 0.26 | 0.35 |
| **Mongar** | 2.61 | 0.94 | 0.64 | 0.36 | 0.50 | 0.44 | 0.28 | 0.35 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Paro** | 2.85 | 0.93 | 0.67 | 0.33 | 0.50 | 0.44 | 0.26 | 0.32 |
| **Pemagatshel** | 2.11 | 0.95 | 0.55 | 0.45 | 0.44 | 0.51 | 0.38 | 0.44 |
| **Phuentsholing** | 1.61 | 0.97 | 0.40 | 0.60 | 0.31 | 0.66 | 0.57 | 0.59 |
| **Punakha** | 2.02 | 0.95 | 0.53 | 0.47 | 0.41 | 0.54 | 0.41 | 0.46 |
| **Samtse** | 1.39 | 0.96 | 0.31 | 0.69 | 0.35 | 0.61 | 0.59 | 0.67 |
| **Sarpang** | 1.50 | 0.97 | 0.35 | 0.65 | 0.35 | 0.62 | 0.57 | 0.64 |
| **SJ** | 1.75 | 0.96 | 0.45 | 0.55 | 0.42 | 0.54 | 0.46 | 0.54 |
| **Tashigang** | 1.87 | 0.93 | 0.50 | 0.50 | 0.46 | 0.47 | 0.39 | 0.48 |
| **Tashi yangtse** | 1.73 | 0.93 | 0.46 | 0.54 | 0.46 | 0.47 | 0.41 | 0.52 |
| **Thimphu** | 2.79 | 0.96 | 0.65 | 0.35 | 0.53 | 0.43 | 0.26 | 0.34 |
| **Trongsa** | 1.69 | 0.91 | 0.46 | 0.54 | 0.46 | 0.45 | 0.40 | 0.51 |
| **Tsirang** | 2.18 | 0.94 | 0.57 | 0.43 | 0.46 | 0.48 | 0.35 | 0.42 |
| **Wangdue** | 2.03 | 0.90 | 0.56 | 0.44 | 0.41 | 0.48 | 0.36 | 0.42 |
| **Zhemgang** | 1.72 | 0.93 | 0.46 | 0.54 | 0.41 | 0.53 | 0.45 | 0.52 |

The Frequency Bias Index (B) for most stations in the southern parts of the country is below 2. The bias was seen to increase with the altitude.

The Probability of Detection (POD), sometimes called Hit rate ranges between the value 0.88 – 0.99 meaning the forecast was able to capture the event of rainfall 88 to 99% respectively. Gasa has the lowest probability of detection with a value of 0.88. The probability of detection of most of the stations in the south was higher compared to the other stations.

The False Alarm Ratio (FAR) was lower for southern and northern stations and higher for western and eastern stations. Lowest FAR value of 0.2 and highest PAG value of 0.8 was indicated for Gasa. On the other hand, Paro had the highest FAR value of 0.67 and the lowest PAG value of 0.33.

The Hanssen & Kuiper's skill (KSS) and Heidke skill score (HSS) for all stations ranges from 0.4 to 0.6 for most of the stations indicating 40 - 60 percent accuracy of the rainfall event forecast. The lowest score of 0.38 was at Lhuentse, whereas phuentsholing has the highest score with 0.66.

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 2021 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 2021. However, it was over-predicted for Haa and under-predicted for Phuentsholing and Punakha. The forecast for surface minimum temperature for 2021 was captured for most of the stations. However, the minimum temperature was over-predicted for Phuentsholing and under-predicted for Paro and Gasa.

The forecast for rainfall events were well captured for 2021 with Probability of Detection (POD) value of 0.88-0.99. Most of the station points indicate decrease in the bias index with the increase in the altitude.

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 $-\infty$ to $+\infty$. 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 $+\infty$. 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 $+\infty$. 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 a 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).

$KSS=TSS=POD-F$

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