Research Journal of Applied Sciences, Engineering and Technology 8(10): 1255-1258, 2014
DOI:10.19026/rjaset.8.1092
ISSN: 2040-7459; e-ISSN: 2040-7467

Submitted: May 04, 2014 Accepted: July 01, 2014 Published: September 15, 2014

Research Article

Verification of Forecasts from High Resolution Numerical Weather Prediction Model

[1]Nagender Aneja and [2]Thomas George
[1]Universiti Brunei Darussalam, Brunei Darussalam
[2]IBM Research, Bangalore, India

Abstract: Assessment of forecast quality is a critical component for weather model development as well as evaluating the impact on weather sensitive business applications such as renewable energy forecasting, agriculture, insurance etc. This study presents forecast quality results of a high resolution numerical weather model deployed for the country of Brunei at Universiti Brunei Darussalam. We present the monthly accuracy and probability of detection scores for precipitation as well as accuracy scores for Relative Humidity (RH) and Dew Point Temperature (DPT) for the year 2013.

Keywords: Accuracy, forecast quality, probability of detection, weather forecast

INTRODUCTION

A regional weather and climate modelling effort was established by Universiti Brunei Darussalam (UBD) in collaboration with International Business Machines (IBM) Corporation to study challenges in forecasting weather and climate for tropical regions. Brunei Darussalam is located at the northern coast of Borneo. The weather in this region is highly uncertain due to tropical dynamics and leads to interesting research problems in the area of climate and weather. The UBD|IBM Centre of Universiti Brunei Darussalam adapted Advanced Research WRF (DTC, 2005) to carry out real time weather forecasting in Brunei. The WRF Model is a next-generation mesoscale numerical weather prediction system for atmospheric research and operational forecasting needs collaboratively developed by a number of national agencies in USA and is currently supported by NCAR.

Many numerical experiments were conducted to set up model domain and configuration. The configuration used for this work is a three-way nested configuration, which includes nests at resolution of 13.5 km covering maritime continent, 4.5 km nest covering all of Borneo and a 1.5 km innermost nest covering Brunei and surrounding areas. To address orographic influence of complex terrain, 45 vertical levels were used in the numerical experiments out of which lowest ten are in the boundary layer. This configuration was placed into operation in November 2011, producing a 48 h numerical weather prediction forecast per day, initialized at 00 UTC using GFS data.

The model is being run every day and a 1.5 km horizontal resolution output is made available to relevant stakeholders via a webpage using Deep Thunder (IBM, 2012). More specifically, we provide maps and animations of humidity, precipitation, wind speed and direction, temperature and water height overlaid over the map of Brunei. The water heights are produced using a flow inundation model that takes the precipitation estimates as input from the weather model. In addition to these, the webpage also provides detailed forecasts at a 10 min temporal resolution for certain locations that are of importance to stakeholders. The main focus of this work is to assess the accuracy and Probability of Detection (POD) of precipitation forecast obtained from the model. We also present Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) for continuous variables such as Dew Point Temperature (DPT) and Relative Humidity (RH). Model Evaluation Toolkit (MET) was used for estimation of forecast quality (DTC, 2007). The results reveal that the model provides an accuracy of about 81% when averaged for all stations, months, different accumulation intervals and thresholds.

Model Evaluation Tools (MET): Model Evaluation Tools (MET) verification package (DTC, 2007) was developed by the National Center for Atmospheric Research (NCAR) Developmental Testbed Center (DTC). MET is a highly-configurable suite of verification tools that can ingest output from the Weather Research and Forecasting (WRF) modelling system. The package includes statistical tools for forecast evaluation, including traditional measures for categorical and continuous variables (e.g., Critical Success Index (CSI) and RMSE).

Corresponding Author: Nagender Aneja, Universiti Brunei Darussalam, Brunei Darussalam

METHODOLOGY

The MET package provides a number of verification measures (Fowler *et al*., 2012) and depending on the forecast variable, a subset of measures is applicable. We treat rainfall as a categorical one and use metrics such as Accuracy (ACC) and Probability of Detection (POD). For calculations we assume that the forecasted event is a dichotomous categorical event and non-probabilistic, i.e., the forecasted event either occurs or does not occur. For example, if the predicted rain was 20 mm and if the observed rain is greater than or equal to 20 mm, then the forecasted event would be categorized as a rain event and similarly a rain of less than 20 mm would be considered as no rain event. We also developed a webpage to visualize the verification results for the high resolution numerical weather prediction system (Aneja and George, 2014). One can select different parameters such as weather station, accumulation interval, threshold for rain/no-rain and day of forecast to compare values with respect to ACC and POD. Our stations include Anggerek Desa, Bandar Seri Begwan, Universiti Brunei Darussalam, Lumpas, Tutong, Sinaut, KB, Sungai Liang, Labi Belait, Sukang, Pekan Bangar, Labu and Brunei International Airport. Predicted values at these stations were compared with actual values taken from Brunei Darussalam Meteorological Service (BDMD, 2014). In order to compare scores, we used different accumulation periods e.g., 1, 3, 6, 12 and 24 h, respectively. In other words for every threshold of rain, one can choose an associated accumulation period for which that threshold is valid. For example, if one categorizes an event as a rain event if 20 mm rain occurred in 3 h of accumulation period, then the total rain in that 3 h period will be used to classify it as rain or no-rain event. Since our numerical weather model provides forecast for every 48 h daily, we evaluated the model skill for multiple time intervals such as 12-36 h, 12-48 h as well as 36-48 h. The first 8-12 h are normally ignored to account for spin up time for the model.

We prepared a contingency table as shown in Table 1 to get values of all the indices.

Based on the contingency table, we can have four possible outcomes:

- Both forecast and observation indicate a rain event (Hits)
- Forecast indicates rain event and observation indicates no-rain event (False Alarm)
- Forecast indicates no-rain and observation indicates rain event (Miss)
- Both forecast and observation indicate no-rain event (Correct Negatives)

Table 1: Contingency table

		Observed		
Forecast		Yes	No	
	Yes	Hits (a)	False alarms (b)	a+b
	No	Misses (c)	Correct non-events (d)	c+d
		a+c	b+d	a+b+c+d

Mathematically, forecast accuracy is (*a*+*d*) / (*a*+*b*+*c*+*d*). Accuracy value can be misleading if there are lots of no-rain events. For example, forecast accuracy may not be useful for predicting low frequency events, i.e., severe thunder storms since there is a strong bias created by the large number of no-rain events. Probability of Detection (POD) or Hit Rate provides information for fraction of observed events that is forecasted correctly. Mathematically, POD is *a*/ (*a*+*c*).

We also computed the errors in estimating DPT and RH. DPT is the temperature at which water vapor in air at constant barometric pressure condenses into liquid water at the same rate at which it evaporates. Dew point temperature is associated with relative humidity. Relative humidity of 100% indicates the air is maximally saturated with water. MAE measures closeness of forecasts to the actual outcomes. The mean absolute error is given by:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|f_i - a_i| = \frac{1}{n}\sum_{i=1}^{n}|e_i|$$

where f_i is predicted value, a_i is actual observed value, e_i is error. Therefore, the mean absolute error is an average of the absolute errors. RMSE is the square root of the mean of the square of all of errors:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(f_i - a_i)^2}$$

RESULTS

Experiments were conducted to determine the accuracy and probability of detection for accumulation intervals of 1, 3, 6, 12 and 24 h; thresholds of 0, 5, 10, 20 mm; forecast duration of 12-36, 36-48 and 0-48 h, respectively. For all the figures in this section, we present the monthly aggregated statistics for 12-36 and 36-48 h from each operational 48 h forecast run made in 2013. Our goal in choosing these two time intervals was to check the accuracy of the last few hours of the forecast period. The first 8-12 h of the forecast are typically ignored to account for spin up time for the weather model.

Figure 1 shows accuracy of precipitation forecasted for 12-36 and 36-48 h and averaged on all 14 stations for 2013 for an accumulation interval of one hour and threshold of 0 mm. We chose a threshold of

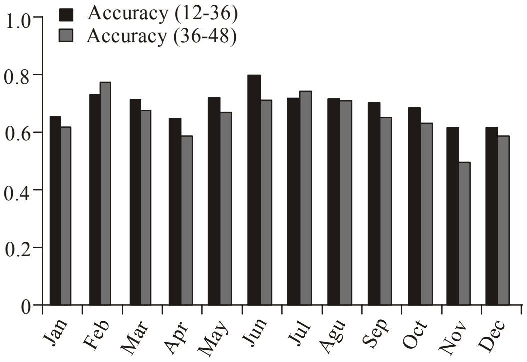

Fig. 1: Accuracy of precipitation averaged for all Brunei stations with hourly accumulation and threshold of 0 mm calculated for 12-36 and 36-48 h

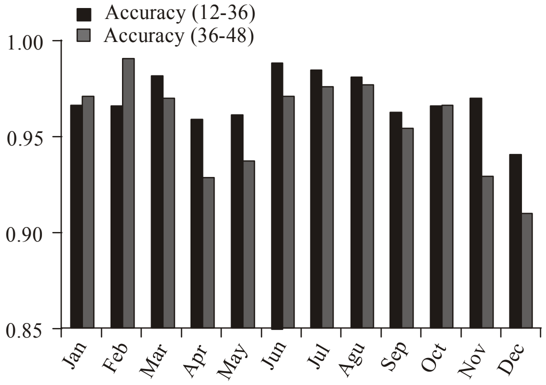

Fig. 2: Accuracy of precipitation averaged for all Brunei stations with hourly accumulation and threshold of 5 mm calculated for 12-36 and 36-48 h

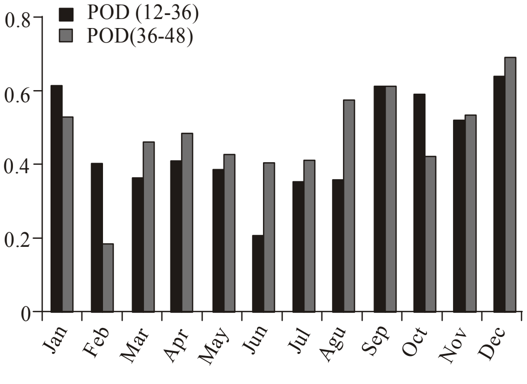

Fig. 3: POD of precipitation averaged for all Brunei stations with hourly accumulation and threshold of 0 mm calculated for 12-36 and 36-48 h

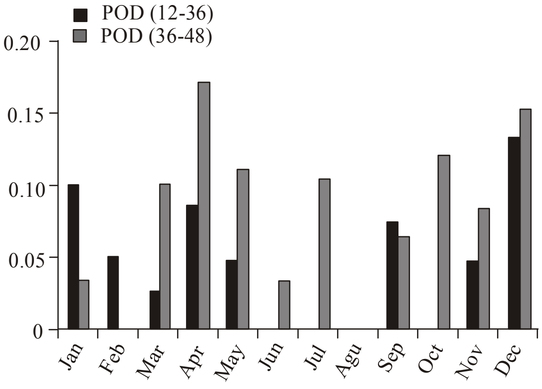

Fig. 4: Accuracy of precipitation averaged for all Brunei stations with hourly accumulation and threshold of 5 mm calculated for 12-36 and 36-48 h

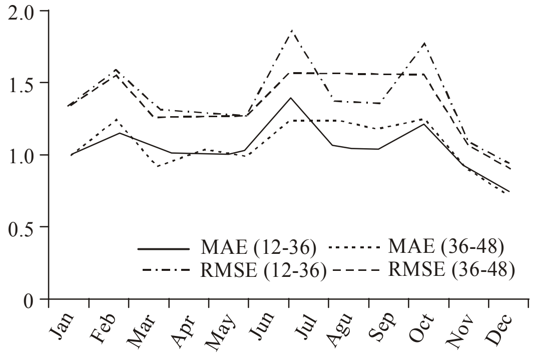

Fig. 5: Mean absolute error and root mean square error for hourly dew point temperature (degree Kelvin) averaged for all Brunei stations

0 mm to categorize rain and no-rain event even in case of small rain. We observe that even for heavy rainfall months such as January and February we have an accuracy of at least 66% for such a low threshold and high temporal resolution. However, if we use a threshold as low as 5 mm as shown in Fig. 2, the accuracy scores jump to above 96% due to the large number of no-rain events. This highlights the inherent problem with relying on the accuracy score alone for evaluating a weather model. The higher threshold results in a significant reduction of the number of rain events.

Figure 3 shows probability of detection of rain events forecasted for 12-36 and 36-48 h and averaged for all Brunei stations with hourly accumulation and threshold of 0 mm. When we use a threshold as low as 5 mm as shown in Fig. 4, we observe a drop in POD scores. The drop in POD scores is due to the large number of no-rain events and number of misses becomes more pronounced in this case. The high resolution weather model is at a disadvantage for point verification of forecasts due to the issue of double penalty. The forecast might have some slight spatial and temporal shift due to its high resolution and since we are comparing only 14 points (with data gaps) out of the possible 283×283 grid points for which forecast is made available, it is not a fair metric. As part of future work, we plan to perform spatial verification of these forecasts using coarse resolution satellite data as well as the high resolution radar data available at a resolution of 2.5 km.

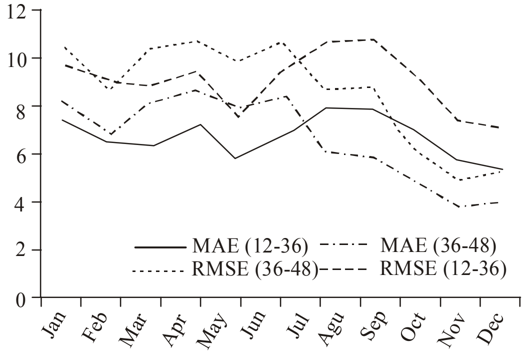

Fig. 6: Mean absolute error and root mean square error for hourly relative humidity (%) averaged for all Brunei stations

Figure 5 shows MAE and RMSE for hourly DPT in degrees Kelvin forecasted for 12-36 and 36-48 h and averaged for all Brunei stations. We observe that the temperature forecasts have a good accuracy as the errors are in the range 0.7-1.9° Kelvin. Figure 6 shows mean absolute error and root mean square error for relative humidity for all 14 stations. The errors are in the range 5-10% for relative humidity for both the time intervals which is considered adequate for most applications.

RECOMMENDATIONS

In future, we plan to conduct sensitivity studies using ensembles over longer durations to improve the accuracy scores for precipitation, temperature, wind speed and relative humidity. In addition to accuracy and probability of detection, we plan to include other scores such as Critical Success Index (CSI), GSS (Gilbert Skill Score), Frequency Bias etc., using a multi-category classification for precipitation. Some of the planned efforts to improve forecast accuracy include data assimilation using radar and satellite data and statistical machine learning to remove spatial and temporal shifts in forecasts.

REFERENCES

Aneja, N. and T. George, 2014. Alpha Testing for High Resolution Numerical Weather Prediction Model (Alpha- TW). Retrieved from: http://202.93.220.97:8080/alpha/.

BDMD, 2014. Brunei Darussalam Meteorological Department, Ministry of Communications, Brunei Darussalam, Jabatan Kajicuaca Brunei Darussalam. Retrieved from: http://www.bruneiweather.com.bn/.

DTC, 2005. The Weather Research and Forecasting Model. Developmental Testbed Center. Retrieved from: http://www.wrf-model.org/index.php.

DTC, 2007. Model Evaluation Tools. Developmental Testbed Center. Retrieved from: http://www.dtcenter.org/met/users/.

Fowler, T.L., T.L. Jensen and B.G. Brown, 2012. Introduction to Forecast Verification. National Center for Atmospheric Research, USA. Retrieved from: http://www.dtcenter.org/met/users/docs/presentations/WRF_Users_2012.pdf.

IBM, 2012. Deep Thunder. Retrieved from: http://www-03.ibm.com/ibm/history/ibm100/us/en/icons/deepthunder/.