Validation and bias correction techniques to improve Numerical Weather Prediction wind speed data

A.G Rangaraj, Krithika.V, Y.Srinath,
K.Boopathi and N Sheelarani
R&D and RDAF
National Institute of Wind Energy (NIWE)
Chennai, India
Email id: rangaraj.niwe@nic.in

Sushant Kumar
National Centre for Medium range Weather Forecasting (NCMRWF)
Noida, India

K.Balaraman
Director General
National Institute of Wind Energy(NIWE)
Chennai, India

***Abstract*—Wind power forecasting is most significant for integration of wind power generation into the Grid. Accuracy of wind prediction is one the biggest challenges faced by researchers around the world as it depends on the Numerical Weather Prediction (NWP) data. The NWP models have evolved over a period of time. However, it is noted that even with various statistical post-processing techniques, there is a void in the accuracy of wind speed forecast with NWP models. This is mainly due to the limitations in the data assimilation process, numerical approximations, parameterization, spatial/temporal resolution and non-availability of actual terrain/ground measurement data. The accuracy of medium term NWP model data output typically day-ahead or week-ahead can be heavily impacted due to variation in localized or unforeseen weather changes in a particular site location. Wind speed and wind direction of NWP model data are the most crucial parameters in the modelling of wind power forecast system.**

In this paper, the monthly, seasonal and yearly NWP wind speed data is analyzed and validated with the measured wind speed data from the Met Mast. In order to effectively identify the offset and biases to correct the Indian NWP wind speed data and hence improved the wind forecasting accuracy, using mathematical, machine learning and hybrid approach. It is acknowledged that various data viz, Indian regional model NWP data, Indian global resolution NWP data and Met Mast data for 13-site locations data were used to carry out this research work.

Keywords- Numerical weather prediction; wind speed; wind power; statistical; machine learning; bias

I. INTRODUCTION

The main fuel of wind turbine generator is wind speed which is mostly affected by large-scale atmospheric conditions and the morphology of the surface landscape [1], this makes wind energy highly intermittent and site-specific. This is one of the major drawbacks for integrating the large scale renewable energy in the Electricity grid. In order to balance the electricity grid with large scale renewable energy, there may be a need for adequate infrastructure to carry out real time ramp up / ramp down of generation or large scale storage system or the need to forecast the power accurately in week ahead / day ahead / intraday basis. Out of these three methods, wind power forecasting would be the most cost effective approach to integrate wind energy in the grid. Thus an accurate forecasting and scheduling system will help wind energy penetration without compromising on the economic aspects and the stability / security of the electric grid.

For carrying out wind power forecasting there are various methods are available and normally any forecaster would select an appropriate method with respect to forecast horizon [2]. As far as day ahead / week ahead wind power forecasting is concerned, an accurate wind speed forecast plays a critical role. Therefore, an accurate wind speed prediction will be used to estimate an accurate wind power forecast. The wind speed forecast shall be obtained from the Numerical weather prediction (NWP) data which is basically an atmospheric model that contains various uncertainties due to practical limitation in the existing technology. Some uncertainties in the wind speed prediction are due to systematic error while others are due to random/unforeseen events. Generally, Statistical/Machine learning algorithms are used to reduce the systematic error in the wind speed forecast. However, due to limitations in the technology it is not possible to correct the random error. In this paper, Indian NWP data set is analyzed with actual measurements collected from different parts of the country and also attempted various existing bias correction methodology to reduce the NWP forecast error.

Based on the initial analysis of NWP and actual measurement data, the average day ahead wind speed forecast absolute error is observed between 2-4m/s and this would impact the accuracy of wind power forecast. Hence, after through data analysis, the suitable bias correction methodology was identified which can be adapted in the operational forecasting system to improve the wind speed forecast.

From the literature survey, the various advantages / disadvantages of different statistical post processing techniques were studied. Most commonly used post processing technique is the Model Output Statistics (MOS) that is used to improve site-specific NWP forecast by determination of the statistical relationship between the weather forecast and observations. [3] As a part of MOS, there are various statistical bias correction techniques to correct the bias in NWP which includes short term bias correction, Diurnal cycle forecast correction, Kalman filter, mean and variance corrected forecast and directional bias forecast [4]. It was also noted that the combined forecast technique viz, Artificial Neural network combined forecast (ANN-COM) and Mean square error-combined forecast (MSECOM) had a better skill score than individual bias correction techniques. These bias techniques were applied on Consortium for small-scale modeling (COSMO) Model output across several locations in Ireland. [4] Similarly, one such hybrid model is Combing Autoregressive Integrated Moving Average (ARIMA) with ANN which resulted in higher accuracy compared to individual models. [7].

Apart from statistical/machine learning models, Physical approach also normally utilized to improve the wind speed forecast. In the physical approach, dynamic downscaling shall be carried out with the help of Linear flow modelling tools or Computational Fluid Dynamics (CFD) tools by modelling the local terrain feature such as roughness, obstacle and orography so that it can convert the mesoscale data to microscale data with higher temporal resolution. However, the major drawback of the physical model is utilization of extensive computation & time to generate forecast and error in local terrain feature modeling would highly impact the forecast accuracy. [5]. Hence, Combination of physical model and statistical model is used in a typical forecasting system to get the advantages of both physical and statistical model. [6].

Based on the above, in this paper, four individual bias techniques and six combinational techniques to bias correct the wind speed of regional and global NWP forecast were discussed in detail. This paper is classified into four major sections and the first section is Introduction. In section II, describes the site and nature of the input data. Section III, explains the methodology adopted to carry out different bias correction in the NWP data and section IV provides the results/outcomes of this study.

II. Model Input

A. *Unified NWP Model*

National Centre for Medium Range Weather Forecasting (NCMRWF), a leading weather forecasting centre provides ‘u’ and ‘v’ components of near surface wind from its global and regional model suite running operationally. These models are based on Unified Model adopted from UK Met office. Global model has a horizontal grid resolution of ~12 km with 70 levels in the vertical reaching 80km height in which 11 levels are within 1 km near to the surface for well representation of the surface features. The initial condition of the model is prepared by hybrid 4D-Var data assimilation (DA) method. The DA system runs four times a day produces four analysis at 00, 06, 12 and 18 UTC. NCMRWF receives global meteorological observations through Global Telecommunication System (GTS) via Regional Telecommunication Hub (RTH) at IMD, New Delhi and satellite observations through internet data services, directly from various satellite data producers such as NOAA, EUMETCAST, ISRO etc. These datasets then undergo quality control tests before preparing the analysis files. A detailed description of NCMRWF DA system can be found in Kumar et.al.[16]. Model produces winds, temperature and other surface parameters at hourly interval. These datasets are re-gridded to a coarser resolution to 0.25° for easy data handling and quicker processing. Bilinear interpolation technique is used for getting data at coarser resolution. Regional model has a grid resolution of ~4km with 80 vertical levels reaching up to 38.5 km and 14 model levels below 1km near the surface. The model contains some parameterized physical processes, including mixed-phase microphysics (based on Wilson and Ballard, 1999), radiation (based on Edwards and Slingo 1996) and land-surface (Best et al., 2011) scheme. Orography is derived from the NASA Shuttle *Radar* Topographic Mission (SRTM) 90 meter digital elevation map. The model produces 10 minutes model outputs for selected variables like surface wind speed, temperature, rainfall, shortwave radiation etc. NCUM-R uses the high-resolution analysis prepared by regional 4D-Var system. In addition to the most of the observations used in the global model Indian Doppler Weather Radar observations are assimilated in the regional DA system.

The various sources of error in the NWP model outputs can be attributed to initial condition error, model parameterizations, different physics packages used and data assimilation [5]. Additionally, for a regional model the lateral boundary condition may also contribute to the errors.

NWP global and regional wind speed data were obtained from NCMRWF from June 2016 to July 2019 with 1-hour temporal resolution (time block) and wind speed and Wind Direction is derived for each time block using (1) and (2) respectively.

$$Wind\ speed = \sqrt{v^2 + u^2}\ m/s \qquad (1)$$

$$Wind\ direction = 57.3 * tan^{-1}\frac{u}{v} + 180° \qquad (2)$$

B. *Measured Wind Speed*

As part of, National wind resource assessment programme initiated by Government of India, NIWE installed wind monitoring masts in all the states of India for estimating the wind potential in the country.

In order to validate the NWP data, authors have selected 13-different sites based on wind regime, terrain condition, data availability, and geographic location as shown in Figure (1). A typical wind monitoring station consists of various meteorological sensors to collect data viz., wind speed, wind directions, temperature and pressure which are captured at different heights. The data logger collects raw data at 2 Hz frequency and the same is averaged to 10 minutes or 15minutes interval [8]. List of wind monitoring stations considered for this study is mentioned in Table 1. The validation was carried out for the global model with 2 heights viz., 10m and 50m, regional model with 50m height using actual measurements and applied 10 different Bias correction techniques to improve the accuracy of NWP model. Since NCMRWF data is 1-hour temporal resolution, the 15minutes / 10minutes measured wind speed was resampled to hourly average data before carrying out validation process.

1) Data cleaning process

Wind monitoring stations (WMS) are locating at remote site, NIWE installed GPRS modem for collecting data through online. In case of non availability / weak GPRS signal, the data is being collected manually from site by replacing the memory card on a monthly basis. Even though Data is transmitted with utmost care yet there can be errors due to malfunctioning of measuring devices or human error, network connectivity and missing data. Hence, before initiating the validation process with WMS data, it is required to identify and clean the erroneous data . The following quality check were performed on the measured data before carrying out the validation process in NWP data.

a. Identification and removal of duplicate timestamps
b. Detection and removal of outliers / anomaly
   i. 0 < Wind speed (m/s) > 25
   ii. Wind direction > 360
   iii. Abnormal values
c. Detection and removal of constant values over a given period i.e., faulty readings due to malfunction of measuring device
d. Visual inspection of the time series data

TABLE 1: LIST OF WIND MONITORING STATIONS

Station ID	State	Type of terrain	Analysis Start date	Analysis End date
1	Tamilnadu	Plain	2018-05-01	2019-05-31
2	Tamilnadu	Plain	2017-07-31	2018-07-31
3	Tamilnadu	Plain	2016-12-24	2017-12-24
4	Tamilnadu	Plain	2017-12-01	2018-12-31
5	Tamilnadu	Plain	2018-03-01	2019-03-31
6	Tamilnadu	Complex	2018-04-24	2019-05-24
7	Tamilnadu	Plain	2017-02-01	2018-02-28
8	Assam	Semi-Complex	2017-08-01	2018-08-31
9	Assam	Semi-Complex	2018-01-01	2019-01-31
10	Madya Pradesh	Plain	2016-08-01	2017-08-18
11	Meghalaya	Complex	2017-11-13	2018-11-13
12	Meghalaya	Complex	2018-01-01	2019-01-31
13	Telangana	Plain	2018-06-01	2019-06-30

C. Determination of window length(WL)

Data analysis was carried out for each station and identified the overall performance of NWP with the help of heat map, box plots, scatter plots and line graphs as shown in Figure (3).

Based on the data analysis, it was observed that the forecast error is not consistent and the same is varying over period of time. Therefore, selection of an appropriate window length will improve the bias correction model. A simulation model was created to run the model with different window length using sliding window technique and optimum window length is selected for each bias correction model. The same process is performed in all the WMS. In the simulation model, window size ranges were considered as 3 - 30days. RMSE error metric is used to select the optimum wind length and model performance. The simulation model would run on daily basis for the analysis start / end date as mentioned in the Table 1 to generate corrected time series NWP data

Figure 1:Geographical location of 13-stations in India

III. METHODOLOGY

A. Simulation Model:

1) Bias correction model

Simulation model has been created for the 4 bias correction model and the details of each model is explained in subsequent section. For each station, the developed simulation model shall use the past x days of raw NWP data / actual measurement data using sliding window technique, as explained above, as training data and estimate the respective model bias dynamically. Then, the model bias will be applied in the day ahead NWP forecast for correcting the NWP dynamically.

2) Combinational model

For each station, the developed simulation model shall use the past x days of bias corrected model output data / actual measurement data using sliding window technique, as explained above, as training data and estimate the weightage of individual bias correction model using combinational model algorithms dynamically. Then, the calculated weights will be applied in the day ahead bias corrected NWP forecast to generate the combination forecast dynamically.

B. Bias Correction Techniques

The measured wind speed, forecasted wind speed and bias corrected wind speed are denoted as WSN, WSF, and BWS respectively. In this paper, the bias correction techniques viz, Short term bias-correction forecast (STB), Diurnal bias correction forecast (DRL), Univariate linear regression (LR) and Directional-Bias correction forecast (DIR) was tested. The brief description of each method is explained below:

1) Short term bias correction (STB).

For each station, the developed simulation model shall estimate the overall model bias as explained above. Then the estimated bias will be applied in the corresponding day ahead wind speed forecast for correction using (3)(4). While applying the necessary bias correction in the NWP, the wind speed values are corrected to zero, in case the corrected wind speed is less than zero. This bias correction method mainly aims to correct the overall bias dynamically in the NWP data.

$$Bias_w = \frac{\sum_{n=1}^{24\times w}(WSN-WSF)}{24\times w} \quad (3)$$

$$NWP^{w}_{Corrected} = Bias_w + WSF \quad (4)$$

where, w = window length (3,4,…30) and n = 1 hour block

2) Diurnal cycle forecast correction (DRL)

Seasons can be classified into three types based on the average wind speed recorded at the site location,

a. Windy Season: High wind speed are recorded during the period May to September

b. Non-Windy Season: Low wind speed are recorded during the period November to March

c. Transition season: Period between windy and non-windy season - October and March

Figure 2(a)and (b) shows the hourly average wind speed measured for a period of 4 years in one of the RE rich states of India based on seasons. A pattern can be observed in the diurnal cycle of all the three seasons, it is evident that the wind speed gradually decreases from 00:00 to 06:00 and gradually increases till it reaches a peak during 15:00 -17:00 i.e decrease in the wind speed during the night and higher wind speed during the day. Hence, irrespective of all the season change in wind speed hour wise was observed. Therefore, under this bias correction model, the hour wise bias in NWP were reduced.

For each station, the developed simulation model shall estimate the hour wise model bias. Then the estimated hour wise bias will be applied in the respective hour of day ahead wind speed forecast for correction using (5)(6). While applying the necessary bias correction in the NWP, the wind speed values are corrected to zero, in case the corrected wind speed is less than zero. This bias correction method mainly aims to correct the hour bias dynamically in the NWP data.

$$Bias_w^h = \frac{\sum_{n=1}^{w}(WSN^h - WSF^h)}{w} \quad (5)$$

$$NWP_w^h = Bias_w^h + WSF^h \quad (6)$$

where, w = window length (3,4,…30) and h = hours (h= 1,2,3..24)

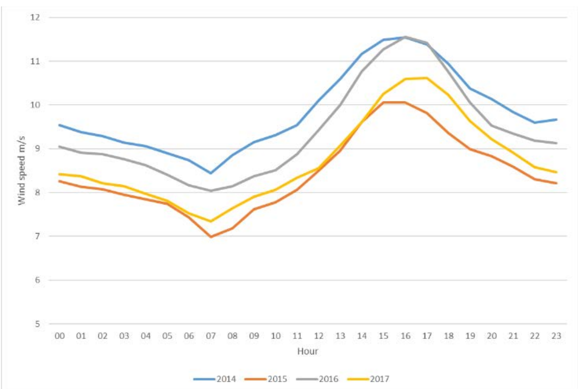

Figure 2 (a):Diurnal pattern during windy season

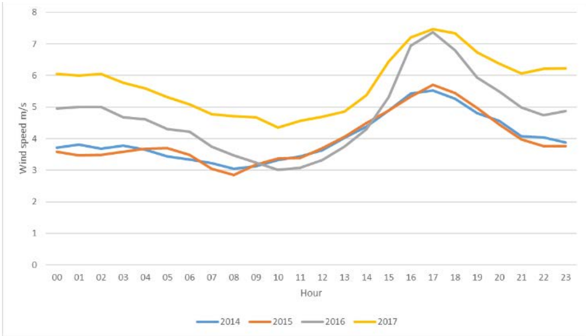

Figure 2 (b):Diurnal pattern during non-windy season

3) Directional-Bias forecast (DIR)

The wind direction provides vital information regarding the change in weather conditions at a particular site. Similarly, the wind farms are designed with consideration of predominant direction of wind speed. Thereby it is suggested to incorporate wind direction parameter into the bias correction model. Typically Wind direction ranges from 0-360°, where 0 is Northern direction.

For each station, the developed simulation model shall estimate the sector wise model bias . Then the estimated sector wise bias will be applied in the respective sector of day ahead wind speed forecast for correction using (8) (7) and overall bias will be used if any of the sector is not available in the day ahead forecast. While applying the necessary bias correction in the NWP, the wind speed values are corrected to zero, in case the corrected wind speed is less than zero. This bias correction method mainly aims to correct the sector wise bias dynamically in the NWP data.

$$Bias_w^s = \frac{\sum_{n=1}^{w}(WSN^s - WSF^s)}{w} \quad (7)$$

$$NWP_w^s = Bias_w^s + WSF^s \quad (8)$$

where, w = window length (3,4,…30) and s = Sector (s= 0-30^0, 30-60^0 ,60-90^0, 90-120^0 ,120-150^0, 150-180^0 , 180-210^0, 210-240^0 ,240-270^0, 270-300^0 , 300-330^0, 330-360^0)

In this method for each window length, The Wind direction data are grouped into 30° sectors. For each wind direction sector, the bias is estimated and overall bias is used for non-availability sectors for the specific window length. Sector wise bias is then applied to the corresponding sector day-ahead forecasted wind speed data.

4) Univariate linear regression (LR)

For each station, the developed simulation model shall estimate the linear relationship (regression coefficients) between forecasted wind speed and actual wind speed using simple linear regression model [4]. Then the estimated regression coefficient applied in the day ahead forecasted wind speed using equation (9). While applying the necessary bias correction in the NWP, the wind speed values are corrected to zero, in case the corrected wind speed is less than zero.

$$BWS = a + (b * WSF) \quad (9)$$

Where, a = constant and b = regression coefficient

C. *Combinational Techniques*

Based on the analysis of all the individual models, it was understood that there is no single method that performs best for all given periods and site locations. So in order to improve the NWP wind speed, all the models should be considered by applying the necessary weights. The weights are estimated with various techniques such as Inverse RMSE weighted average (IRMSE), Gradient Boost (GB), linear regression (LR), Random Forest (RF), K-Nearest Neighbor (KNN) and extreme gradient boost (XGB). The same moving window technique was adapted to dynamically change the weight age for different weather condition.

Each combination model has a varied mechanism to calculate the weights for each variable based on the training data (10) [12].

$$BWS = f_1W_1 + f_2W_3 + \cdots f_nW_n \tag{10}$$

Where, f = Model output of individual models and W = weights

1) *Inverse RMSE weighted average (IRMSE)*

For each station, the developed simulation model shall estimate the RMSE for individual bias correction model. Inverse of RMSE shall be considered corresponding weight for the individual bias correction model using (11) (12). Then the estimated weights will be applied in the day ahead bias corrected forecasted wind speed using (12).

$$W = \frac{RMSE^{-1}}{\sum_{f=1}^{n} RMSE^{-1}} \tag{11}$$

$$BWS = \frac{(RMSE_{f1}^{-1} * f_1)+(RMSE_{f2}^{-1} * f_2)+\cdots+(RMSE_{fn}^{-1} * f_n)}{\sum_{f=1}^{n} RMSE^{-1}} \tag{12}$$

Where, f_x = bias corrected wind speed for individual bias technique models

2) *Multivariate Linear Regression (MLR)*

For each station, the developed simulation model shall estimate weight by using multivariate linear regression method, weight for each individual bias correction models were estimated i.e., all the individual bias correction models are considered as independent variables. The calculated regression coefficients will be applied as the corresponding weight of the individual bias correction model. Then the estimated weights will be applied in the day ahead bias corrected forecasted wind speed by using (8).

$$BWS = a + b_1f_1 + b_2f_2 + \cdots + b_nf_n \tag{13}$$

Where, f = individual model output and b = regression coefficient.

3) *Combinational – Random Forecast model (RF)*

It is an ensemble model that combines multiple decision trees. Tree bagging algorithm trains multiple trees based on random sample data points from the dataset. Bagging algorithm trains multiple trees based on random sample features from the dataset. A random forest combines both the techniques and the aggregation is performed by averaging the outputs of all the trees. The number of trees in the forest (n estimator) is set to 100 and the function to measure the quality of split is set to mean squared error (MSE) [13] For each station, the developed simulation model shall estimate the weights for individual bias correction model using Random Forecast algorithm. Then the estimated weights will be applied in the day ahead bias corrected forecasted wind speed

4) *Combinational – Gradient Boost (C-GB)*

Gradient boosting is a boosting algorithm that works on the principle of optimization in the function space i.e stage wise additive expansions of Gaussian mixture density are used to find the optimal model. [14]. Number of boosting stages is set to 100, learning rate is 0.1 and least squares regression as loss function. For each station, the developed simulation model shall estimate weights for individual bias correction model using Gradient boost algorithm. Then the estimated weights will be applied in the day ahead bias corrected forecasted wind speed

5) *Combinational – Extreme Gradient Boost (C-XGB)*

It's a machine learning algorithm that supports various objective functions, including regression, classification and ranking problems, which produces a prediction model in the form of ensemble of decisions trees. Model is constructed similar to other boosting methods, and it generalizes them by allowing the optimization of an arbitrary differentiable loss function. It's capability to do parallel computation on a single machine makes it highly robust and fast. For each station, the developed simulation model shall estimate the weights for individual bias correction model using Extreme Gradient Boost algorithm. Then the estimated weights will be applied in the day ahead bias corrected forecasted wind speed

6) *Combinational – K-Nearest neighbour (C-KNN)*

K-nearest neighbors is technique that averages the target value associated with it's nearest neighbor based on distance metric [15] The number of neighbors (n_estimator) is set to 5 and the left size is set to 30. For each station, the developed simulation model shall estimate the weights for individual bias correction model using K Nearest Neighbour algorithm. Then the estimated weights will be applied in the day ahead bias corrected forecasted wind speed

IV RESULTS AND DISCUSSION

The 10 bias correction techniques were applied for different moving window for 13 stations mentioned in table 1. In this paper, the result of one typical station is given in Figure 3. As per Figure 3, the similar pattern was observed in all the analyzed stations.

Figure 3[a-c] represents the heat map of hour wise raw NWP MAE of 10m (global resolution), 50m (global resolution) and 50m(regional resolution). It is observed that June month is having higher error in all the models. However, 10m height global resolution model has a higher accuracy as compared to 50m (global and regional resolution). This is because at 10m height NWP provider used many measurement points for data assimilation process.

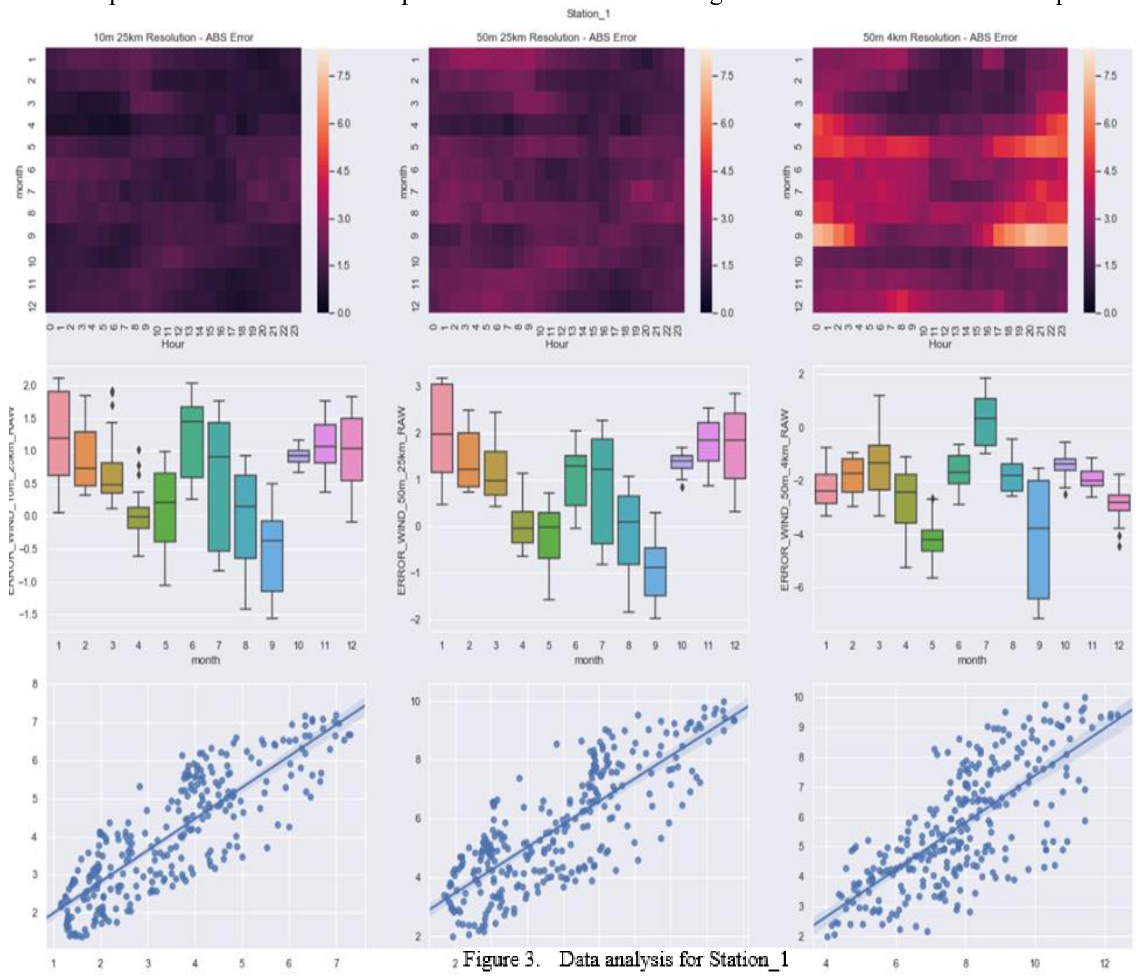

Figure 3. Data analysis for Station_1

It is observed that the MAE is higher and scattered throughout the year for higher resolution NWP model whereas the MAE is lower and concentrated around particular months/hours in the case of lower resolution NWP. The boxplots in Figure 3 indicate the distribution of raw NWP error on a monthly basis. The forecast shows higher bandwidth/range of error during the windy season and minimum during the transition period for all the resolutions. Similarly, the scatterplot shows the correlation between the forecasted wind speed and measured wind speed for all the resolutions. For each bias correction model, overall RMSE is calculated for a typical station. Based on the calculation, the best model and its corresponding window length is identified and summarized in table 2. Similar exerciser was carried out for all the thirteen stations.

Based on the detailed analysis, it can be observed that all the ten bias correction models proposed in this paper has improved the accuracy of RAW NWP in all the 13 stations as shown in Figure 4(a),4(b) and 4(c). From the table 3, it is evident that minimum reduction of RMSE for the 4km resolution (50m) is 9%, maximum reduction of RMSE is 55% and an average reduction of RMSE is 30% upon usage of best bias correction model in all the 13 stations. In case of Multiple Linear Regression combination (MLR) model with 30days window length the same is 8.7%, 54.6% and 29% respectively.

The minimum reduction of RMSE for the 25km resolution (50m) is 10%, maximum reduction of RMSE is 54% and an average reduction of RMSE is 29% upon usage of best bias correction model in all the 13 stations. In case of Multiple Linear Regression combination (MLR) model with 30days window length the same is 7%, 54% and 27% respectively.

Similarly, the minimum reduction of RMSE for the 25km resolution (10m) is 12%, maximum reduction of RMSE is 70% and an average reduction of RMSE is 34% upon usage of best bias correction model in all the 13 stations. In case of Multiple Linear Regression combination (MLR) model with 30days window length the same is 11%, 70% and 33%respectively.

Based on the above, Multiple linear regression (MLR) combination model with 30days window length is always producing optimum result for different weather / terrain conditions and moreover the difference between the best bias correction method RMSE and MLR's RMSE of is less than 0.1 m/s on an average. The comparison of best bias correction model (BM) RMSE and MLR 30days window length RMSE is shown in Figure 4(a)(b)(c).

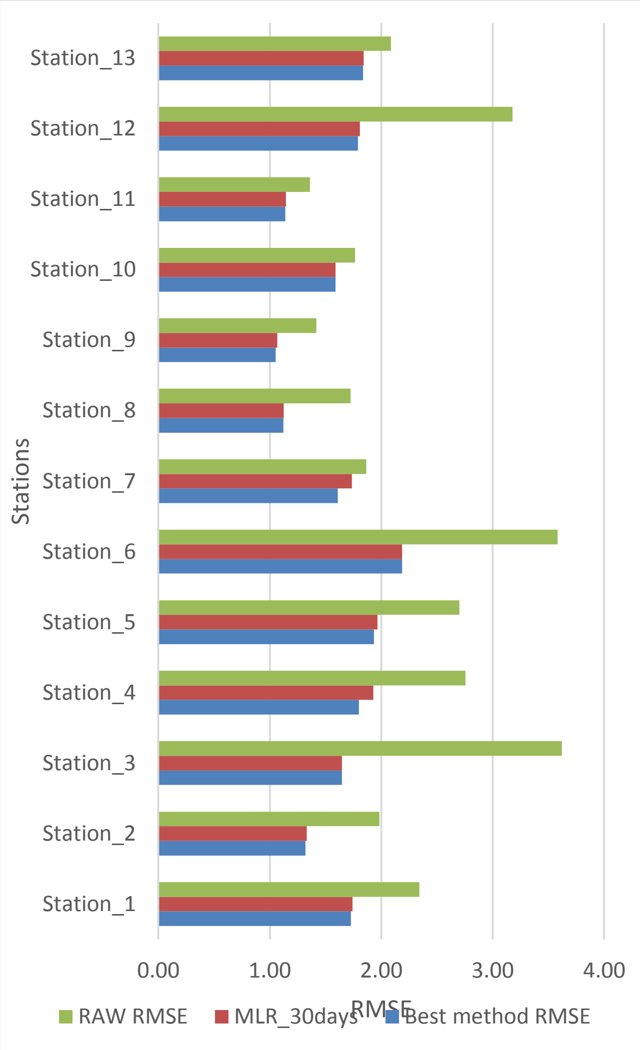

Figure 4 (a):Comparison of best models with raw 25km resolution (50m)

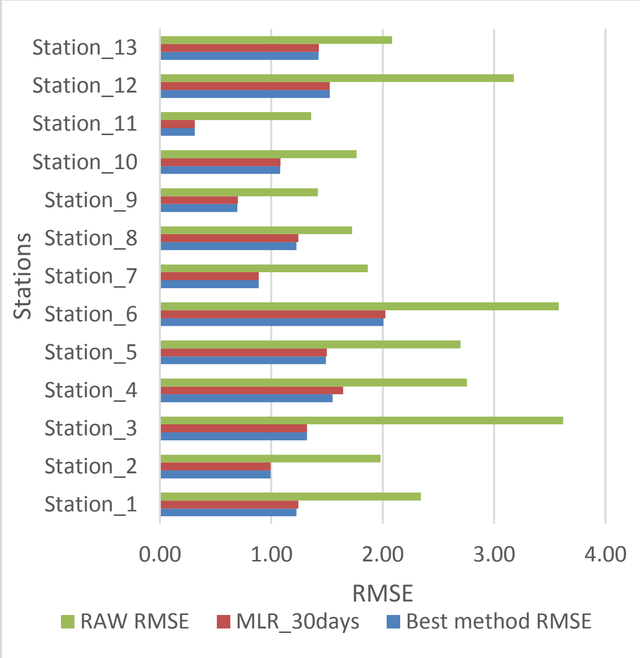

Figure 4 (b):Comparison of best models with raw 25km resolution (10m)

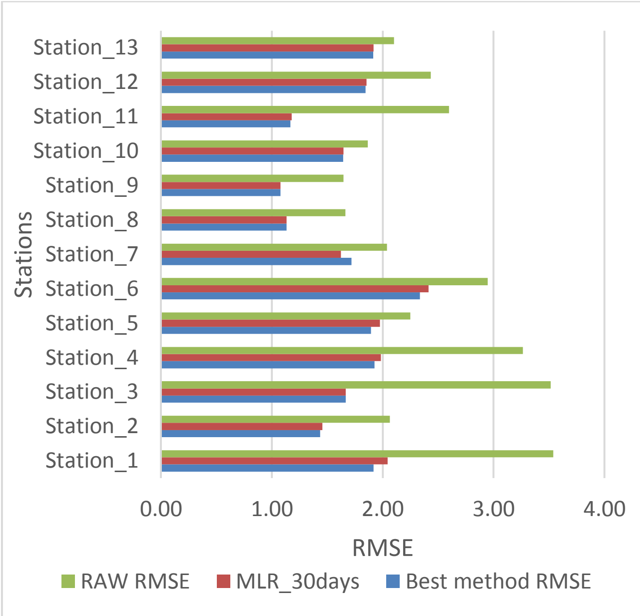

Figure 4 (c):Comparison of best models with raw 4km resolution(50m)

TABLE 2: COMPARISION OF ALL THE METHODS FOR STATION_1

4km resolution (50m)			25km resolution (50m)			25km resolution (10m)		
Methods	**RMSE**	**WL**	**Methods**	**RMSE**	**WL**	**Methods**	**RMSE**	**WL**
RAW	3.86	Average	RAW	2.4	Average	RAW	1.62	Average
STB	2.37	3	STB	1.91	3	STB	1.36	3
DIR	2.4	3	DIR	1.82	3	DIR	1.29	3
DRL	2.3	3	DRL	1.94	3	DRL	1.36	3
IRMSE	2.05	3	**IRMSE**	**1.73**	**3**	**IRMSE**	**1.23**	**3**
LR	**1.92**	**3**	LR	1.78	3	LR	1.28	3
MLR	1.97	16	MLR	1.74	26	MLR	1.24	27
C-GB	2.07	15	C-GB	1.79	29	C-GB	1.27	30
C-KNN	1.94	3	C-KNN	1.82	3	C-KNN	1.3	24
C-XGB	2.05	22	C-XGB	1.78	30	C-XGB	1.25	30
C-RF	1.97	3	C-RF	1.77	WL	C-RF	1.3	24

TABLE 3: COMPARISION OF BEST METHODS WITH MLR-30DAYS FOR ALL THE STATIONS

	4km resolution (50m)				25k resolution (10m)				25km resolution (50m)			
Station ID	BM	WL	IMP% - BM	IMP% - MLR_30D	BM	WL	IMP % - BM	IMP% - MLR_30D	BM	WL	IMP % - BM	IMP% - MLR_30D
1	LR	3	45.8%	42.2%	IRMSE	3	22.3%	21.1%	IRMSE	3	26.2%	25.7%
2	LR	5	30.5%	29.6%	MLR	30	35.7%	35.7%	LR	24	33.5%	32.9%
3	MLR	22	52.7%	52.6%	MLR	30	60.2%	60.2%	MLR	30	54.5%	54.5%
4	LR	3	41.0%	39.2%	LR	3	30.7%	26.4%	LR	3	34.8%	30.0%
5	LR	3	15.8%	12.2%	C-RF	30	33.4%	32.9%	C-XGB	30	28.3%	27.2%
6	IRMSE	3	20.7%	18.1%	IRMSE	3	45.6%	45.1%	MLR	30	39.0%	39.0%
7	LR	7	15.6%	20.3%	MLR	29	36.8%	36.7%	MLR	22	13.7%	6.9%
8	MLR	30	32.0%	32.0%	IRMSE	5	15.3%	14.2%	MLR	29	34.9%	34.9%
9	LR	30	34.5%	34.5%	MLR	25	29.3%	28.7%	LR	25	25.7%	24.8%
10	MLR	27	11.9%	11.7%	MLR	20	11.7%	11.6%	MLR	30	10.0%	10.0%
11	LR	22	55.0%	54.6%	MLR	30	70.0%	70.0%	LR	23	16.3%	15.8%
12	MLR	29	24.2%	23.8%	MLR	30	43.4%	43.4%	MLR	23	43.6%	43.1%
13	MLR	28	9.0%	8.7%	MLR	27	11.6%	11.4%	MLR	21	12.0%	11.7%

BM=best model, WL= window length, IMP%,= Percentage improvement

V CONCLUSION

In this paper, validation of global model with two heights 50m and 10m and regional model with 50m height with actual measurements for 13 different locations in India were carried out. From the data analysis and literature survey, 10 different bias correction techniques were selected and identified the optimum window length for each bias correction models. Based on detailed analysis, all the 10 bias correction techniques are improving the accuracy of the NWP models. It is also observed that for each site different bias forecast methods with different window length are performing better. However, multivariate linear regression (MLR) combinational model with 30days window length is always performing at optimum level in all considered stations and average reduction of RMSE of 30% is noted.

FUTURE SCOPE

Hyper parameter turning of machine learning algorithms can be carried out for further research. Also, the effect of bias corrected wind speed in wind power forecasting can be carried out for future research.

ACKNOWLEDGMENT

The author acknowledge National Centre for Medium Range Weather Forecasting (NCMRWF) for providing the numerical weather prediction data to carry out the research work. The authors gratefully acknowledge various NIWE officials viz., Dr.Rajesh Katyal (DDG & DH), Mr. Haribhaskar (DDT), Mr.Vinod Kumar (JE) for their support in providing the necessary MET MAST data. Authors would like to thank project assistants of NIWE / UKFS officials viz., Mr.Shaktivel T, Ms. GJ Janani, Ms.Rakhi K, Ms.Aishwarya S, Mr.Sasi Kumar and Ms.Shraddha D, for their help in collection and segregation of METMAST data. Authors would also like to thank Ms. S.M. Revathy, Project Engineer NIWE for providing valuable inputs about the NWP models.And also the authors express sincere thanks to all the engineers in NIWE who have installed met masts and collected data.

REFERENCES

[1] Wu, Y., & Hong, J. (2007). A literature review of wind forecasting technology in the world. 2007 IEEE Lausanne Power Tech, 504-509

[2] S. Soman, Saurabh & Zareipour, Hamidreza & Malik, O.P. & Mandal, Paras. (2010). A review of wind power and wind speed forecasting methods with different time horizons. North American Power Symposium 2010, NAPS 2010. 1 - 8. 10.1109/NAPS.2010.5619586. Glahn, H.R. and D.A. Lowry, 1972: The Use of Model Output Statistics (MOS) in Objective Weather Forecasting. J. Appl. Meteor., 11, 1203–1211,https://doi.org/10.1175/1520-0450(1972)011<1203:TUOMOS>2.0.CO;2

[3] Glahn, H.R. and D.A. Lowry, 1972: The Use of Model Output Statistics (MOS) in Objective Weather Forecasting. J. Appl. Meteor., 11, 1203–1211,https://doi.org/10.1175/1520-0450(1972)011<1203:TUOMOS>2.0.CO;2

[4] Sweeney, C. P., Lynch, P., & Nolan, P. (2011). Reducing errors of wind speed forecasts by an optimal combination of post-processing methods. Meteorological Applications, 20(1), 32–40. doi:10.1002/met.294

[5] AG.Rangaraj,K.Boopathi,Sushant.K:Operational challenges in wind power forecasting and scheduling in National conference on renewable energy technologies and its Integration with grid, 2019(Proceedings).

[6] Lei, M., Shiyan, L., Chuanwen, J., Hongling, L., & Yan, Z. (2009). A review on the forecasting of wind speed and generated power. Renewable and Sustainable Energy Reviews, 13(4), 915–920.doi:10.1016/j.rser.2008.02.002

[7] Cadenas, E., & Rivera, W. (2010). Wind speed forecasting in three different regions of Mexico, using a hybrid ARIMA–ANN model. Renewable Energy, 35(12), 2732–2738. doi:10.1016/j.renene.2010.04.022

[8] National Institute of wind energy (www.niwe.res.in)

[9] The Weather Prediction (www.theweatherprediction.com)

[10] Schneider, A., Hommel, G., & Blettner, M. (2010). Linear Regression Analysis. Deutsches Aerzteblatt Online.doi:10.3238/arztebl.2010.0776

[11] Rokach L. (2005) Ensemble Methods for Classifiers. In: Maimon O., Rokach L. (eds) Data Mining and Knowledge Discovery Handbook. Springer, Boston, MA

[12] Sweeney, C., & Lynch, P. (2011). Adaptive post-processing of short-term wind forecasts for energy applications. Wind Energy, 14(3), 317–325.doi:10.1002/we.420

[13] Sci-kit learn ([https://scikit-learn.org)

[14] Rusheng Hu, Xiaolong Li and Yunxin Zhao, "Gradient Boosting Learning of Hidden Markov Models," 2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings, Toulouse, 2006, pp. I-I. doi: 10.1109/ICASSP.2006.1660233

[15] Martínez, F., Frías, M. P., Pérez, M. D., & Rivera, A. J. (2017). A methodology for applying k-nearest neighbor to time series forecasting. Artificial Intelligence Review.doi:10.1007/s10462-017-9593-z

[16] Sumit Kumar, A. Jayakumar, M.T.Bushair, Buddhi Prakash J., Gibies George, Abhishek Lodh, S. Indira Rani, Saji Mohandas, John P. George and E. N. Rajagopal (2018). Implementation of New High Resolution NCUM Analysis-Forecast System in Mihir HPCS