Long-term surface wind-speed forecasting for monsoon period at station scales over INDIA: Implementing ML and Statistical models

Rupanjan Banerjee, Somnath Baidya Roy

Centre for Atmospheric Science, IIT Delhi | Email: *rupanjanbanerjeemay1998@gmail.com*

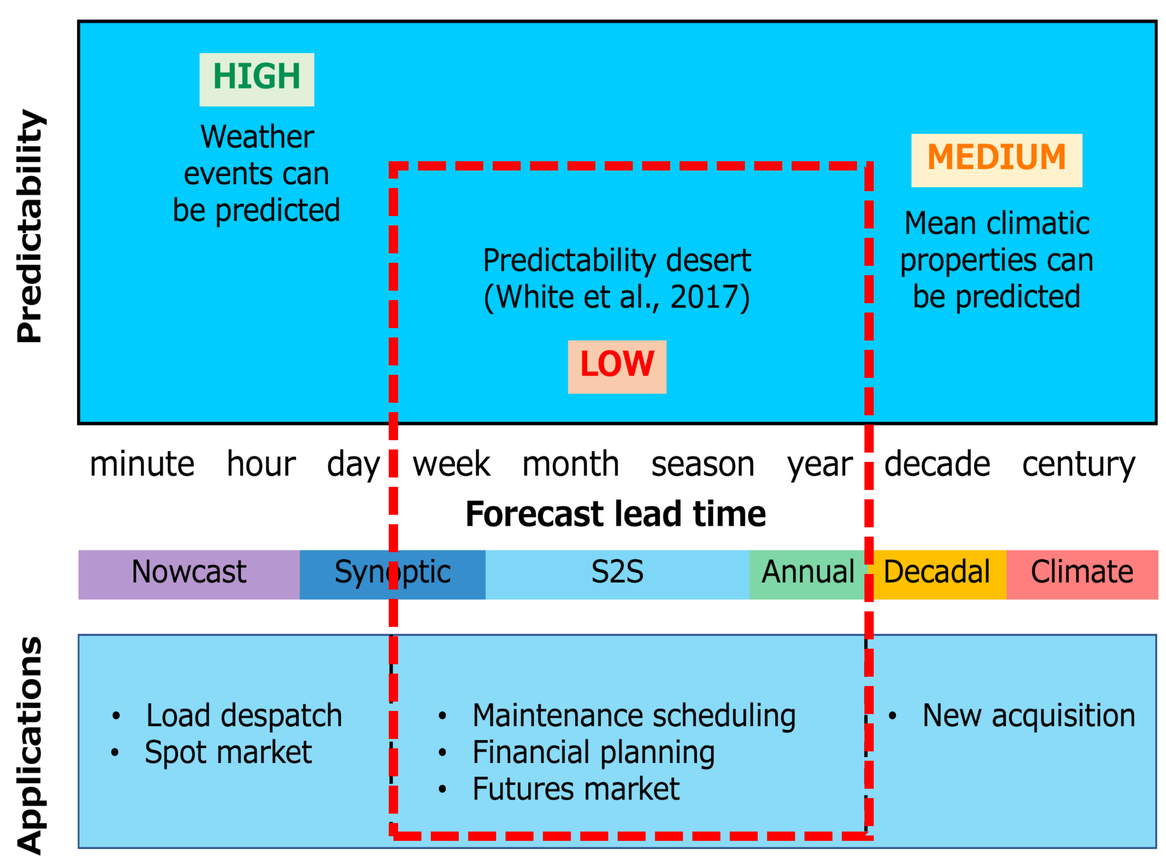
1. Background

The world is gradually shifting towards in using clean and sustainable energy. Currently, India is the fourth largest wind energy producing country, also has the capacity to produce almost 1000 GW of offshore wind energy, which could be a major contributor to India's 50-50 energy goal. Long-term surface wind speed forecasts with 3-4 months of lead times are essential for planning and operations by the wind energy industry. The overall goal of our study is to evaluate and enhance the capability of the IITM CFSv2 model to forecast the summer monsoon (June-September) wind speeds over India at seasonal scales as a part of the Monsoon Mission III program. The model runs were conducted in hindcast mode for the period 1981-2016.

Surface wind speed forecasting at seasonal scale is still in its early stages. Because operational forecasts are not available, the wind industry is forced to rely on climatological averages that cannot capture year-to-year variability. This study can help alleviate this problem and help the industry for operational and planning activities.

Wind speed forecasting challenges :

- A lot of effort is going into S2S forecasting. Agencies like ECMWF, NCEP, CMA, JMA etc. are providing experimental S2S forecasts. But the quality of these forecasts are not good (Orlov et al. 2020, Das and Baidya Roy, 2021).
- Long-term forecasts beyond the synoptic scale are not readily available in India. So, right now the industry has no option but to rely on climatology for many wind farm operations.



2. Broad Goal & Objective

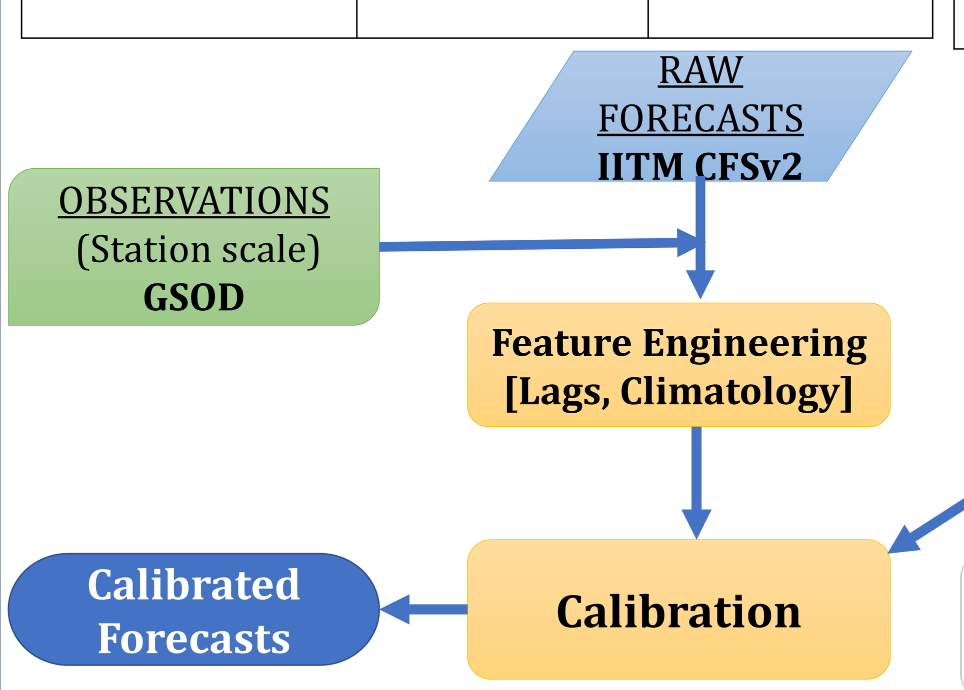
- As part of **Monsoon Mission III** extended project, the broad goal of this work is to evaluate and improve the forecasts from the IITM-CFSv2 model.
- The specific objective is to evaluate and improve the JJAS 10m wind speed forecasts with 3-4 months of lead time at selected stations by calibrating ML models to surpass climatology.

3. Data and Methodology

IITM CFSv2 Model data details:		
Data Used	1) U-component (10m) 2) V-Component (10m) 3) Surface Temperature 4) Upward short-wave flux 5) Downward short-wave 6) SST	
Period	1981 – 2017	
Resolution	0.31° * 0.31°	
Initial Condition	February	March
Forecast Period	Mar 1 – Nov 30	Apr 1 – Dec 31

Global Surface Summary of Day (GSOD) data:	
Data Used	1) Wind Speed (10m) 2) Daily Avg. Temperature 3) Max and Min Temperature 4) Mean Station Pressure 5) Mean Sea Level Pressure
Resolution	Station Scale

SST Data:	
Source:	NOAA OI SST
Regions:	Bay of Bengal, Arabian Sea, IOD, Nino 3.4, Nino 4



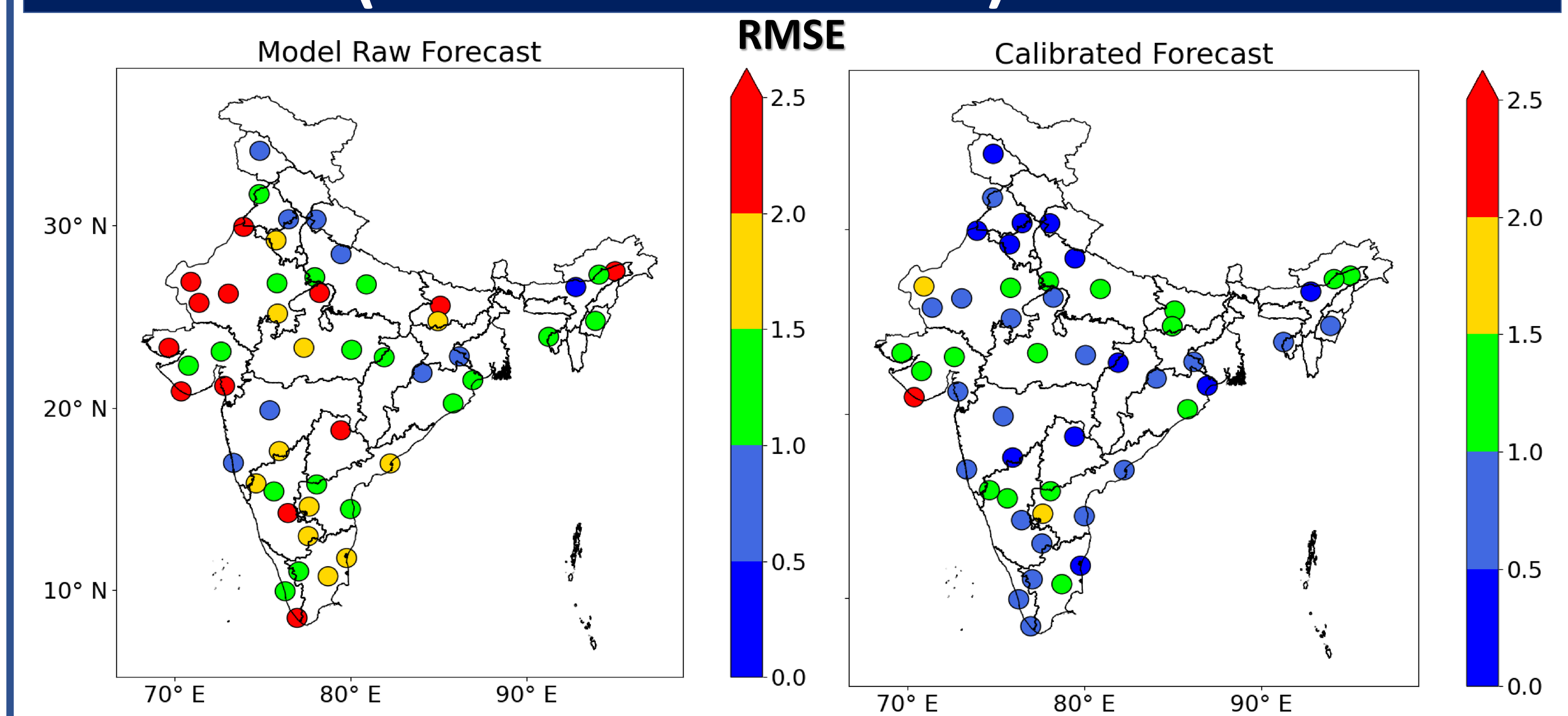
AI-ML Models Used

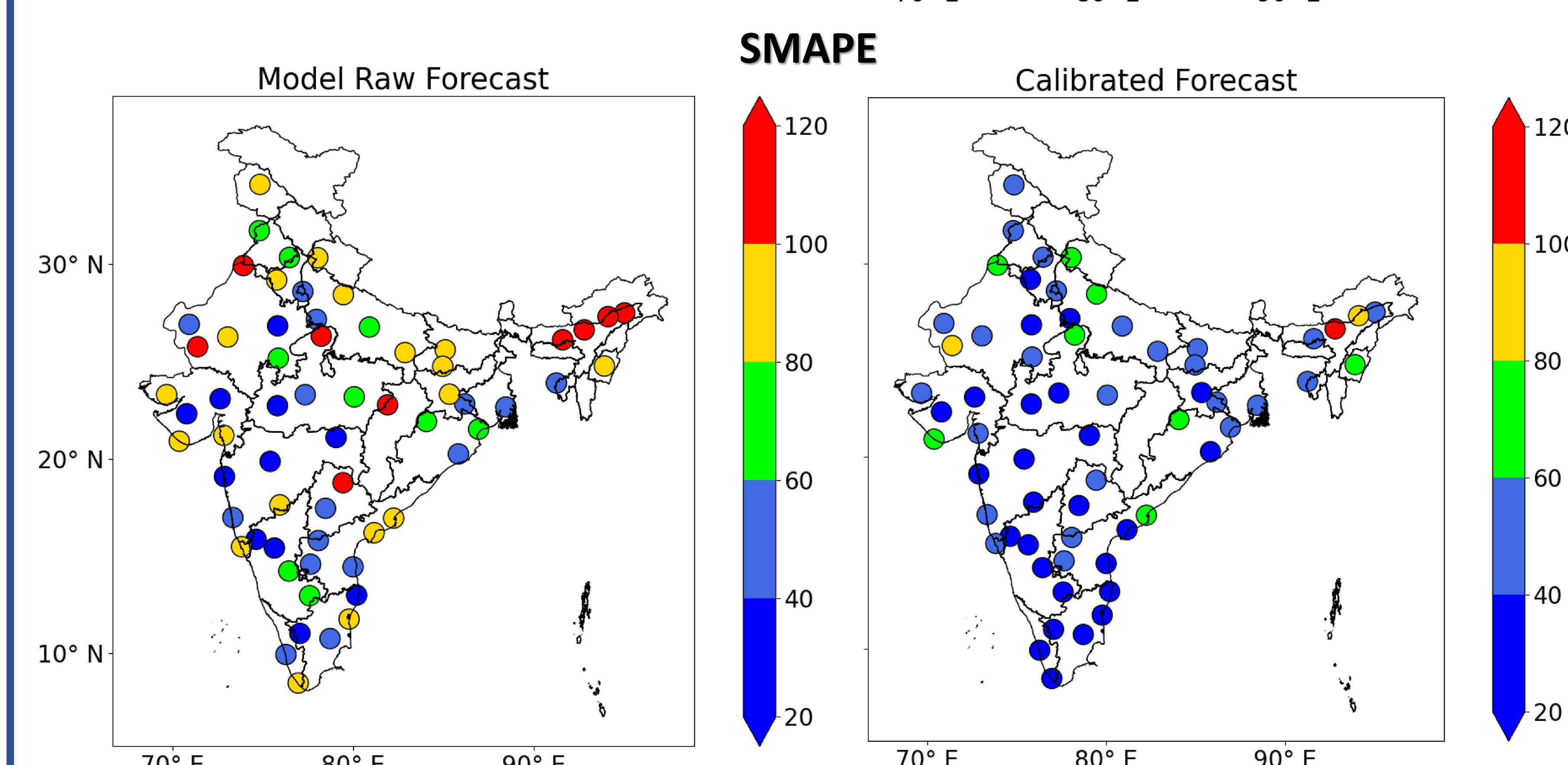
Conv1D-LSTM, Bidir-LSTM, GRU, Optimized Random Forest, Optimized XGBoost, Multi-Linear Regression

$$SMAPE = \frac{1}{n} \times \sum \frac{|\,forecast\ value - actual\ value\,|}{(|\,actual\ value\,| + |\,forecast\ value\,|)/2}$$

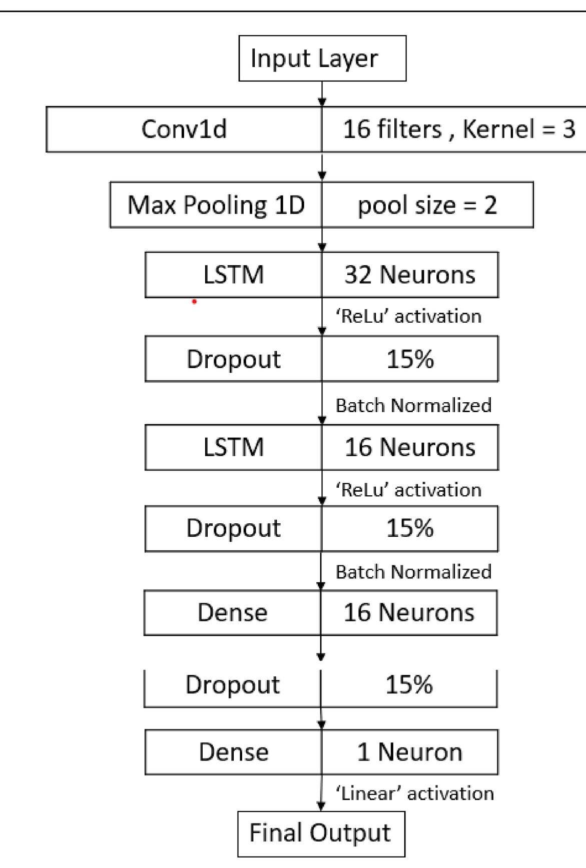
Training Period	Validation Period	Testing Period
1990 – 2010	2011 – 2013	2014 – 2016

5. Results (JJAS of 2014 – 2016)

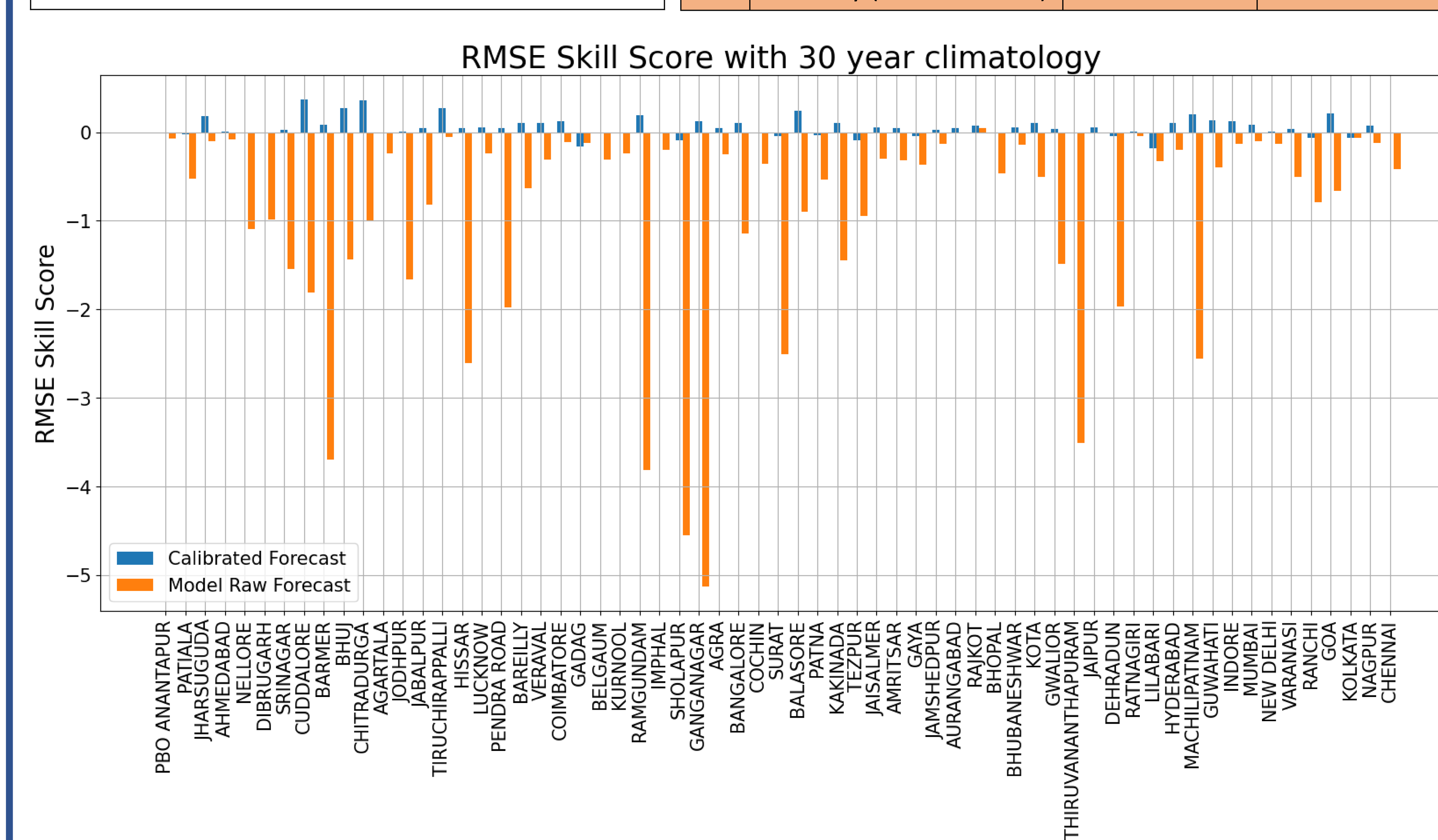




Best Model Architecture



	Model Name	No. of stations with positive RMSESS (Out of 63)	Performance Percentage
1.	Multi Linear Regression	39	~61%
2.	Optimized Random Forest (With Bayesian Optimization)	26	~40%
3.	Optimized XGBoost (With Bayesian Optimization)	27	~42%
4.	Bi-directional Long Short-Term Memory (Bidir-LSTM)	13	~20%
5.	Gated Recurrent Unit (GRU)	32	~50%
6.	Conv1D Long-Short Term Memory (Conv1D-LSTM)	45	~71%



6. Conclusions/Summary

- High RMSE values in wind speed at the stations located in southern part and west region of India are observed with IITM CFSv2 prediction.
- After calibration with AI-ML models, RMSE and SMAPE have improved significantly for all stations along with other statistical metrics.
- SMAPE has been improved for all the stations and for both FEB and MAR IC. 52 out of 63 stations are showing SMAPE within the range 20 – 60%.
- 65 -70 % of stations are showing positive skill score with best calibrated model (Conv1D – LSTM) in both FEB IC and MAR IC where NWP model is showing highly negative skill for all the stations.