Application of Machine Learning for Atmospheric Characterization Using University of Calcutta ST Radar

Arkadev Kundu
Institute of Radio Physics and Electronics,
University of Calcutta, Kolkata, West Bengal, India 700009
[arkadevkundu@gmail.com]

Abstract: An active phased-array Stratosphere-Troposphere (ST) VHF Doppler Radar operating at 53 MHz is being installed at the University of Calcutta's Ionosphere Field Station (22.94°N, 88.51°E, and 34°N geomagnetic latitude) in eastern India, adjacent to the northern Bay of Bengal. It is designed to measure very strong atmospheric winds, especially during the tropical cyclones that frequently hit this region.In this paper, to create machine learning solutions that are customized to the requirements of weather and climate modelling.With this and other radars in India,models of various atmospheric parameters may be developed using measured data over a period of time as training dataset. Applying this model, it can predict the values of the atmospheric parameters for a period outside the training interval. So, the model outputs may help override huge infrastructure and resources necessary for establishing radars.

Keywords – Doppler.Training dataset. Machine learning

1.0 Introduction

The tropical lower atmosphere i.e., troposphere and lower stratosphere holds the key for understanding and modelling of the atmospheric processes. Convection in the tropical troposphere drives atmospheric disturbances ranging from the small-scale turbulence to the planetary scale equatorial waves. A proper understanding of these processes is essential to quantify and model their effects on the atmosphere, in general, and on climate change.

The location of the present station at Kolkata (Ionosphere Field Station) is unique for the lower atmospheric studies as it lies in the transitional zone between the tropics and the subtropical region. This region is characterized by pre-monsoon convective rainfall and Nor'westers which have a profound influence on the day-to-day lives of the common people.

The Radar technology in atmospheric research was started to study the tropospheric radio propagation. Late Professor Sisir Kumar Mitra, founder of the Institute of Radio Physics and Electronics, was one of the pioneers who suggested use of radars for atmospheric research in the 1930s. Atmospheric radars derive information on the atmospheric phenomena by making use of the variations on the frequency, phase, amplitude and polarisation parameters of radio waves which are transmitted from the radar systems, backscattered by the atmosphere and received back by the radar system again. The elements which are used for the atmospheric study are rain, snow, hail and have the wavelength of a few centimetres. A high-power VHF backscatter radar can be used to study atmospheric dynamics up to a height of around 100 km.The University of Calcutta ST Radar is a high sensitivity, high resolution, pulse coded, phase coherent radar operating in the low VHF band (typically around 53 MHz) with an average output aperture product of around$3 \times 10^8 Wm^2$. The dynamic properties of the atmospheric wind, waves, turbulenceand atmospheric stability throughout the atmosphere can be measured by the analysis of the scattered signals.

An active phased-array ST radar is being installed by the University of Calcutta at Haringhata, which is about 50 kilometres northeast of Kolkata, a major city in eastern India, on the eastern bank of the Hooghly River. Notably, it is directly over the Equatorial Ionization Anomaly and the Tropic of Cancer (EIA). It is on the Tropic of Cancer and just below the Equatorial Ionization Anomaly (EIA).

Due to the Kalbaisakhi (Nor'wester) and tropical cyclones, the area around Kolkata is unique in that it has a tropical climate with an average annual rainfall of about 1600 mm. This area is also known for deep convection, thunderstorm and lightning activity. Although part of the Indo-Gangetic Plain (IGP), the phenomenology of this part of the IGP differs significantly from other parts due to atmospheric processes governed by Bay of Bengal. In addition, this area is designated as a "wetland" and is certified as a "wetland of international importance" by the "Ramsar Convention". It also has the world's largest sewage-fed aquaculture and acts as a strong carbon sink.because of all these factors Haringhata an important location to study atmospheric dynamics and coupling processes up to the ionosphere over a tropical region.

The ST Radars provide information on atmospheric parameters in the troposphere and lower stratosphere over a wide range of spatial scales. The most important and unique capability of the ST radars is the measurement of vertical wind component with a high degree of temporal and altitude resolutions (typically ~30sec and ~150m). This unique capability gives the ST radars an enormous advantage over the conventional wind measurement techniques (radiosondes).

The absence of high-quality data that can be used for training is one of the main obstacles for Hard AI in short- and medium-range timeframe weather forecastingUsing conventional models to train machine learning systems could increase the effectiveness of high-performance computing.However, the quality of thepredictions would not be improved.There is scope for prediction improvements if the model data could be generated at a higher spatial resolution than what is currently possible to run during an operational forecasting window. However, the computation power required to generate this training dataset would be huge. Consistent and high-quality data are necessary for training from observations. The University of Calcutta ST Radar could fulfil that requirement by providing high resolution data. Also, it is very important that successful development of a model could substantially reduce the cost of operational forecasts for severe weather conditions like thunderstorms and tropical cyclones. Hence the major motivation is to predict the atmospheric Doppler measured by the ST Radars which may be used to calculate the various components of atmospheric winds, which are the direct deliverables and parameter for issuing weather forecasts.

Temperature measurements of the atmosphere reveal a more detailed vertical structure. The atmosphere can be divided into four zones on the basis of thermal characteristics. These zones are Troposphere, Stratosphere, Mesosphere and Thermosphere as shown in Figure 1.

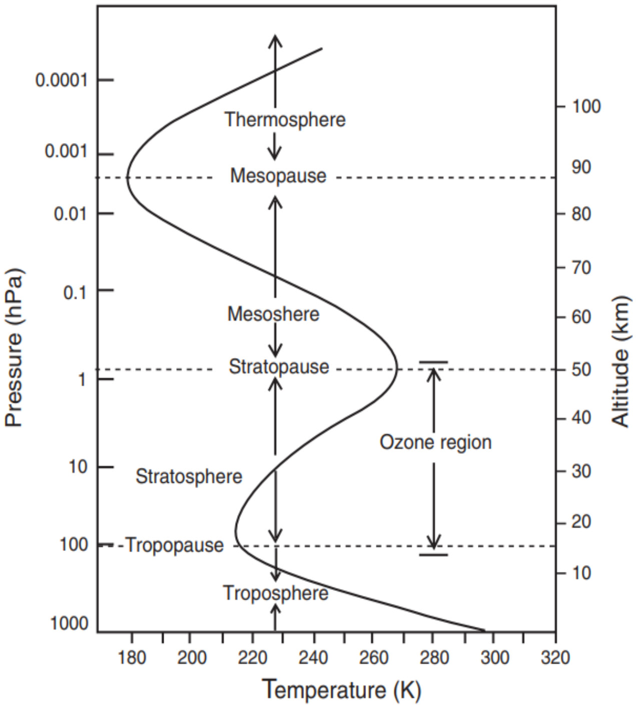
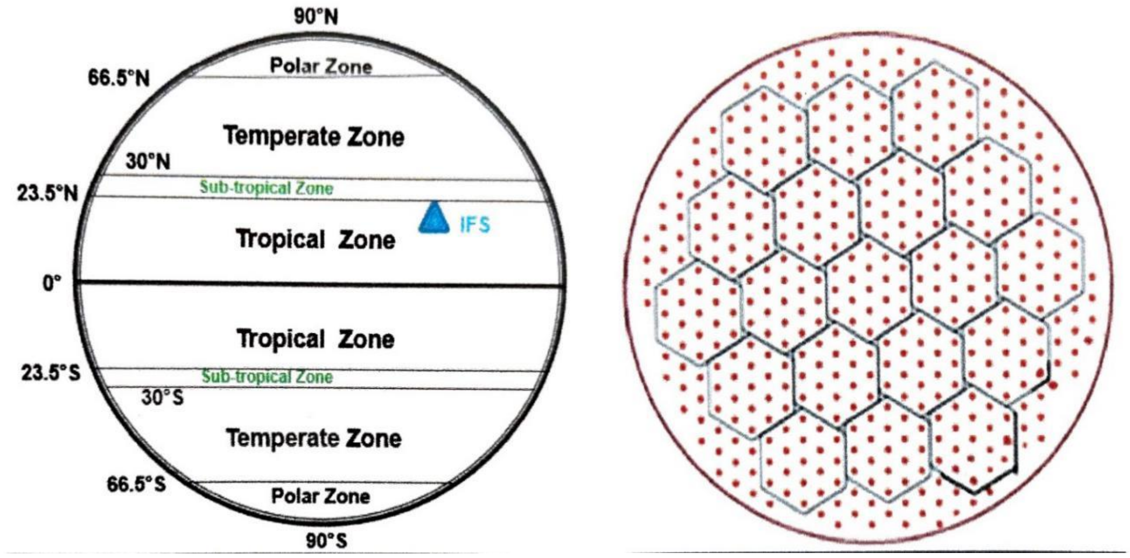


Figure 1. Vertical thermal structure of Earth's atmosphere up to 120 km (Adapted from G. Brasseur and S. Solomon 1984)

The ST (Stratosphere-Troposphere) radar of University of Calcutta has been placed at IFS (Ionospheric Field Station) at Haringhata (Latitude 22.94°North, Longitude 88.51°East) in the eastern part of Indiaadjoining northern Bay of Bengal.This ST radar is located in the sensitive tropical to subtropical geophysical region near the northern crest of the Equatorial Ionization Anomaly (EIA) in the longitudinal sector of India as shown in Figure 2.



(a) (b)

Figure 2. (a) Geographical location of ST Radar, (b) ST Radar main Array Diagram

Table 1- IFS, Haringhata 53 MHz Radar Specifications

PARAMETER	***SPECIFICATION***
Location	Ionospheric Field Stationat Haringhata (Latitude 22.94°North, Longitude 88.51°East)
Frequency	53 MHz
Number of Yagi-Uda antenna	475
Antenna aperture	~7,000 m^2
Antenna Gain	34.1 dB
Antenna Beam width	3.6^o
Peak power	2 kW per TR module, 950 kW (approx. 1MW)
Maximum duty ratio	5%
Scan range	Scanning up to 30^o from zenith direction
Pulse width	0.5 to 128 μs (coded/un-coded)
Maximum no. of coherent integrations	1024
Maximum no. of range bins	512
Maximum number of FFT points	4096
Time Resolution	5-15 minutes for full profile
Receiver	25 super-heterodyne receivers
Receiver bandwidth	8 MHz
Receiver dynamic range	80 dB

ThisST radar at IFS, Haringhata has the capabilities of probing the lower atmosphere up to 22 km and the upper atmosphere up to 600 km. The most important and unique capability of the ST radar is the measurement of the vertical wind component with a high degree of temporal and altitude resolutions. Thus, ST radars are more advantageous over the conventional wind measurement techniques using radiosondes for this unique capability.

ST Radar transmitted the frequency to the atmosphere. There are several particles present, such as dust and water vapour. The transmitting frequency collided with those particles and returned to ST Radar, which received back scatter power with shifting frequency. The frequency shift from the centre transmission frequency is caused by a change in refractive index. That is the source of the Doppler frequency.

1.1 Co-operative Research Program using ST/MST Radars

Presently, there are a total of 5 numbers of ST/MST radars in India. These are: CUSAT ST Radar (205MHz) at Cochin/Kochi (10.0° N, 76.3° E); MST Radar (53MHz) at Gadanki (13.5° N, 79.2° E); CU ST Radar (53MHz) at Haringhata (23.0° N, 88.6° E); Gauhati University ST Radar (212.5MHz) at Guwahati (26. 2° N, 91.7° E) and ARIES ST Radar (206.5MHz) at Nainital (29.4° N, 79.5° E). Currently, there are efforts to develop coordinated research programs using these 5 radars.

Figure 3. shows the 19-element sub array radar set up at IFS, Haringhata. This Pilot array is hexagonal in shape.

ST Radar Pilot Sub-array at IFS, Haringhata

Figure 3.IFS Haringhata ST Radar Pilot Sub-array

2.0 AI/ML model and Data

It is now possible to classifying weather and climate prediction into different timescales because to recent developments in deep neural networks, related machine learning techniques, and statistical forecasting.It is clear that instead of depending on machine learning solutions that are primarily created for other application areas like image recognition, the community has to build machine learning solutions that are specifically customized to the needs of weather and climate modelling.It would be necessary for such customized solutions to be able to impose physical understanding within the solution architecture.This is required to create machine learning tools with the least amount of complexity possible to optimize efficiency during training and inference. Additionally, machine learning solutions for weather and climate applications must be able to cope with changes in dynamic regimes due to climate change, so they must be able to train outside of available training systems for past weather. In ideal but unlikely scenarios, the solution, in addition to interpolation, can accurately estimate where standard deep learning would succeed.

With these radars, models of various atmospheric parameters may be developed using measured data over a period of time as training dataset. Applying this model, it can predict the values of the atmospheric parameters for a period outside the training interval. There are extensive use of AI/ML based techniques to implement this model. Refinement of the model outputs may help override huge infrastructure and resources necessary for establishing radars.

In this paper the objective is Develop a AI/ML based model to predict various atmospheric parameters using data recorded by the CU ST Radar.

- Prediction of Doppler at various height with respect to time.

- Measurement of relative error between observed and predicted Doppler.
- **2.1 Facebook Prophet**

Prophet is an open-source forecasting tool available for Python and R made to help developers researchers and analysts more efficiently set goals and allocate resources over weeks or years.

It is fully automatic so that non-experts in data forecasting can quickly make high quality forecasts but also allows for hand tuning so that experts can improve their results by adding specialized knowledge. Prophet additive regression model performs best with time series that exhibit strong seasonal effects and historical data from multiple seasons. It can automatically detect changes in trends by selecting change points from the data and model yearly seasonal components using the Fourier series, however itis also robust to missing data and trend changes and generally handles outliers well.It has the following features:

- **Accurate and fast:**Prophet is utilised in a variety of Facebook applications to generate reliable forecasts for planning and goal setting.In the majority of cases, it has been found to perform better than any other approach. By fittingmodels in Stan and get predictions in just seconds.
- **Completely automated:**With no manual effort, it is possible to obtain a reasonable forecast from messy data.Prophet is robust to outliers, missing data, and dramatic changes in the time series.
- **Tunable forecasts:**The Prophet method provides numerous options for users to customise and adjust forecasts.By adding domain knowledge, human-interpretable parameters can be used to improve predictions.
- **Available in Python or R:**The Prophet procedure has been implemented in R and Python, but the underlying Stan code for fitting is the same in both.It is acceptable to use whatever language one is comfortable with to get forecasts.

2.2 Training dataset for ML model

- The dataset is created for the ML model from the ASCII files.
- Period of data is used to train the ML model: 6 months (January, 2020 to June, 2020)
- Doppler measured in height with respect to time: 1.35km to 10.2km.
- Resolution of the height: 0.15km.
- Radar operations per day: around 6 hours.

Figure 4. shows the presently operational main ST Radar at IFS Haringhata.

Figure 4. Operational main ST Radar at IFS Haringhata

2.3 Brief explanation of the Data and Methodology

Six months of Doppler datafromJanuary 2020 to June 2020have been used as a training dataset in this machine learning model of the University of Calcutta's Pilot Sub-Array at the Ionosphere Field Station (IFS), Haringhata. Five beams with an off-zenith angle of 15° and directions of Doppler data east, west, zenith, north, and south were used.

- Cross validation: Cross validation is the process of training learners using one set of data and testing it using another set. Prophet includes features for time series cross validation to measure forecast error using past data.
- Hyperparameter tuning: Hyperparameter tuning is the process of choosing values for a model's parameters that maximize the model's accuracy.The model's hyperparameters, such as ***changepoint prior scale*** **and** ***seasonality prior scale***, can be tuned via cross-validation.

a) ***changepointpriorscale:*** This is probably the most effective parameter. Determines trend flexibility, especially how much the trend changes at trend changepoints. If it is too small, the trend will be underfitted and the variance that should be modeled on the trend change will be treated by the noise term instead. If it is too large, the trend will overfit, and in extreme cases, it may end up with a trend that will capture yearly seasonality. The default value of 0.05 works for many time series, but it can be adjusted. A range of [0.001 - 0.5] is probably good.

b) ***seasonalitypriorscale:*** This parameter controls seasonal flexibility.Similarly, a large value allows seasonality to large variations, and a small value reduces the magnitude of seasonality. The default value is 10, which applies basically no regularization. It should probably be tuned within the range of [0.01 - 10].

The complete process flowchart is shown in Figure 5.

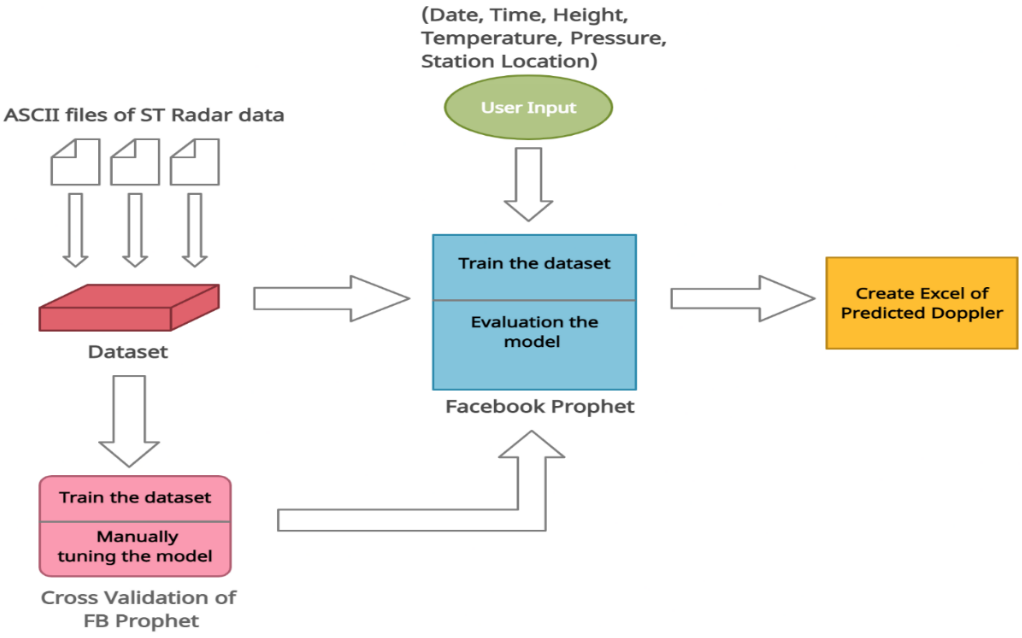
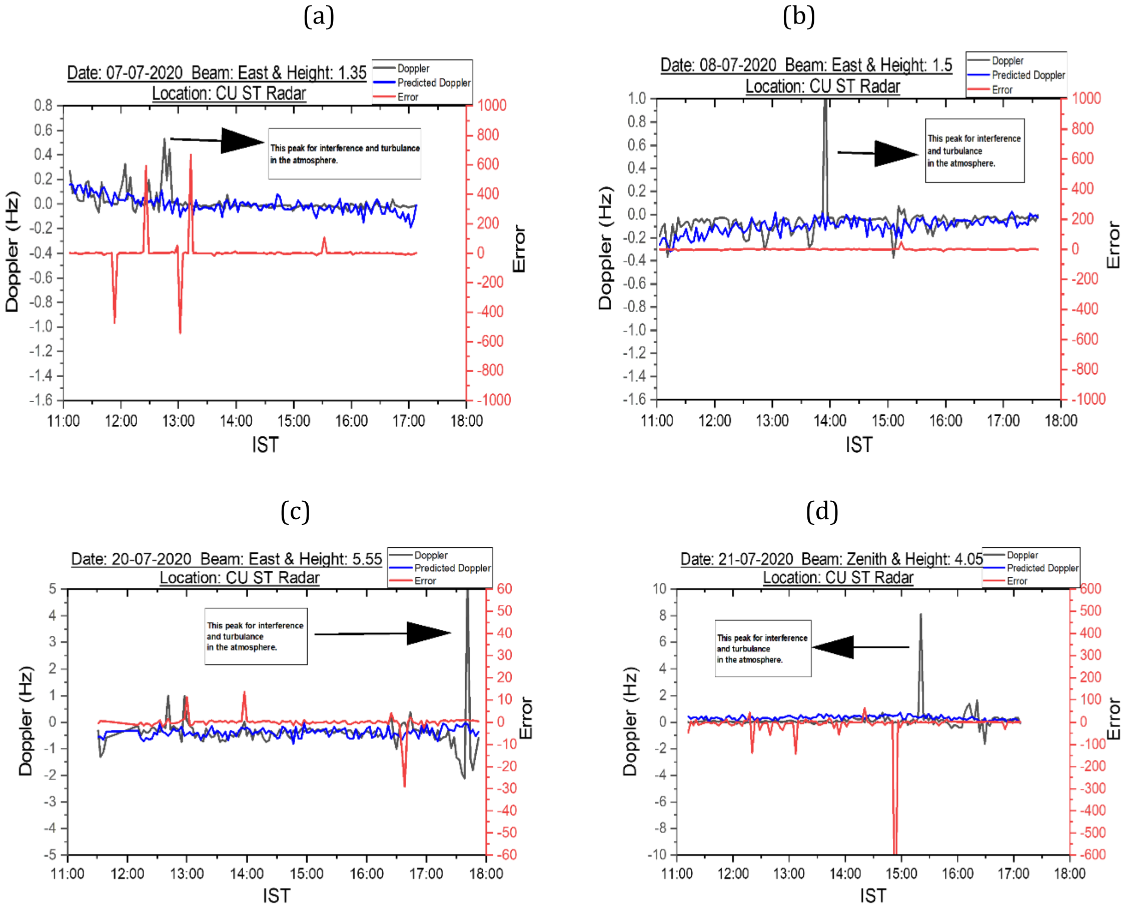


Figure 5. Flowchart of the complete process

3.0 Results

In this machine learning model using the Pilot Sub-Array at Ionosphere Field Station (IFS), Haringhata, of the University of Calcutta 6 months (January, 2020 to June, 2020) of Doppler data have been used. Five beams with an off-zenith angle of 15° and directions of Doppler data east, west, zenith, north, and south were used.The basic objective is predict the Doppler atmospheric parameterso that the radial velocities could be calculated. Then after applying the formula, we get the three components of the wind vector (Vertical, Zonal and Meridional), wind speed and wind direction.This approach, if successful, could significantly reduce the huge infrastructure cost required for building ST Radars.

Figure6shows the comparison between the actual measured Doppler and predicted Doppler.



(a) (b) (c) (d)

Figure 6. Comparison between the actual measured Doppler and predicted Doppler of (a) East beam of 1.35 km height on 7th July 2020, (b) East beam of 1.5 km height on 8th July 2020, (c) East beam of 5.55 km height on 20th July 2020, (d) Zenith beam of 4.05 km height on 21st July 2020.

Figure 7 shows the distribution of Relative error with respect to time.

(a) (b)

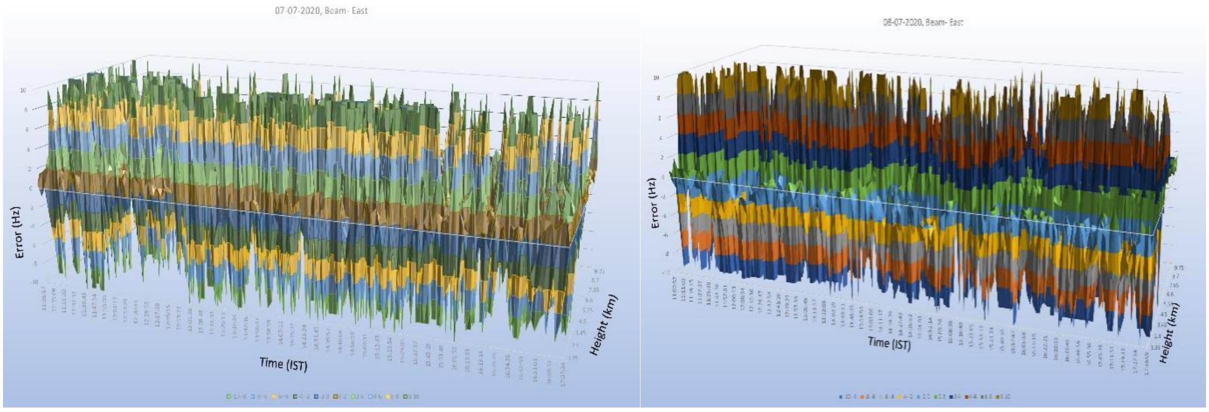


Figure 7. Distribution of Relative error with respect to time (a) East beam on 7th July 2020, (b) East beam on 8th July 2020.

Figure 8 shows the distribution of Relative error with respect to height

(a)

(b)

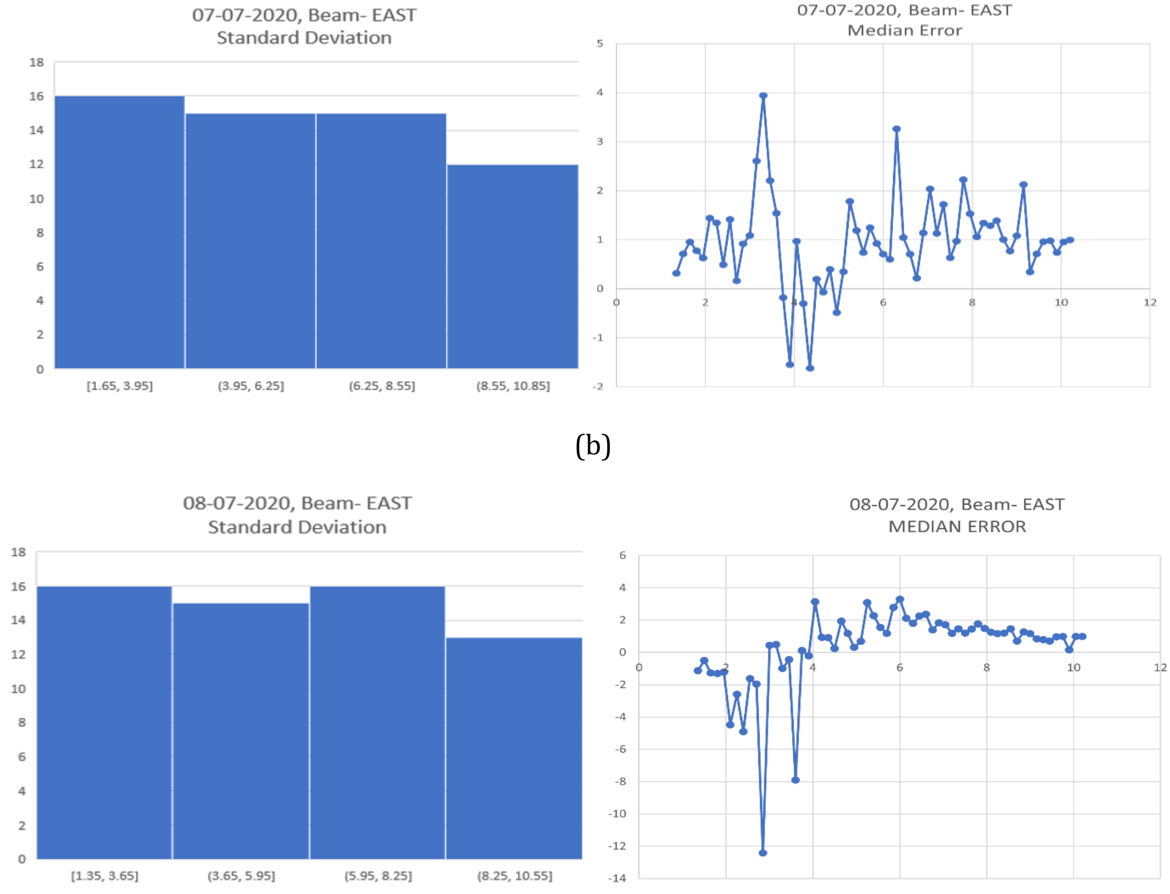


Figure 8. Distribution of Relative error with respect to height (a) East beam on 7th July 2020, (b) East beam on 8th July 2020.

4.0 Conclusion and Future work

- Development of this model is not complete as evident from the errors between the measured and the model values. In future, data covering a longer period may be used to train the ML model. The predicted Doppler obtained from this model will then be validated with actual observations to test the performance of this model.
- The differences occurring between measured data and model data could be arising because of lack of inclusion of time (for diurnal variations), day of year (for seasonal effects) and point of scattering (for spatial variations) in the present version. These needs to be included.
- The error values between measured and model Doppler values have been truncated to ±10Hz to eliminate the spikes in Doppler attributed to interference and flight of airplanes over the radar area.
- The standard deviation and median values of the relative error are less at higher height (>7km) than at lower heights.

5.0 References

- Mohanakumar, K., Straosphere- 1 roposphere Interaction: Au Introduction, Science & Business Media, 2010. Springer https://doi.org/10.1007/978-1-4020- 8217-7
- P.Nandakumar, Debyendu Jana, Sunilkumar S V, Satheesh Chandran.P.R Validation of Wind Measurements From a 53 MHz ST Radar Pilot Array Located at University of Calcutta With Collocated Radiosonde Launches, March 2022. DOI:10.1029/2020RS007246.
- UCAR Center for Science Education article on Layers of Earth's Atmosphere https://scied.ucar.edu/learning-zone/atmosphere/layers-earths-atmosphere
- Geeks for Geeks article on MLMachine Learning - GeeksforGeeks
- Facebook Prophet websiteProphet | Forecasting at scale. (facebook.github.io)
- Holton J.R., Haynes P.H., McIntyre E.M., Douglass. R.A., Rood B.R. and Pfister L., Stratosphere-Tropospherexchange, 1995 https://acd-ext.gsfc.nasa.gov/People/Douglass/95RG02097.pdf
- M. I. Skolnik, Introduction to Radar Systems, 3rd Edition, McGraw-Hill, New York, 2001.
- Machine Learning https://certes.co.uk/types-of-artificial-intelligence-adetailed-guide