URSI-RCRS 2024, ARIES/GEHU, Bhimtal, India, 22 - 25 October, 2024

Advanced Turbulence Characterization using XGBoost Algorithm

Ahana K K(1, *), Amal Joy (1), Satheesan K(1,2), Avinash Paul(1), and Ajil Kottayil(2)
(1) Department of Atmospheric Sciences, Cochin university of Science and Technology, Pallimukku, 682016
(2) Advanced Centre for Atmospheric Radar Research, CUSAT, Kalamassery, 682022

Abstract

A machine learning approach was used to better understand turbulent kinetic energy dissipation rate (ε) estimated from S-T radar located at Kochi, in relation to background conditions. A supervised machine learning model was utilized, where ε was the predictand, and level, day of the year, time of day, horizontal wind speed, vertical wind, and wind direction were the predictors. Overall the model yielded a correlation coefficient of 0.886 between the modeled and training datasets. An explainable Artificial Intelligence (XAI) technique was used to understand the relation between the predictand and predictors which revealed that level was the most influential variable. Regarding background wind conditions, horizontal wind speed was identified as the primary factor contributing to turbulence, with increased wind speed correlating with higher turbulence. On the other hand, lower vertical wind speed values were found to be associated with increased turbulence.

1. Introduction

Turbulence plays a crucial role in atmospheric dynamics, influencing weather patterns, pollutant dispersion, and aviation safety. Accurate estimation and analysis of turbulence are essential for improving weather prediction models and understanding atmospheric processes. Traditional methods of estimating turbulence often rely on in-situ measurements, which can be sparse and limited in spatial and temporal coverage. In recent years, remote sensing technologies, such as wind profiler radars, have emerged as powerful tools for observing atmospheric turbulence. The Stratosphere-Troposphere (S-T) wind profiler radar, in particular, offers high-resolution vertical profiles of wind speed and direction, providing valuable data for turbulence estimation. However, interpreting and analyzing the vast amounts of data generated by these radars can be challenging.

Machine learning techniques have shown great promise in handling large datasets and uncovering complex patterns within them. In this study, we leverage the capabilities of the XGBoost machine learning algorithm to analyze turbulence estimates derived from an S-T wind profiler radar. By integrating machine learning with radar observations, we aim to gain deeper insights into the factors influencing turbulence and improve the accuracy of turbulence estimation. Furthermore, we employ SHAP (SHapley Additive exPlanations) analysis to interpret the contributions of each predictor to the model's output. This approach allows us to identify the most influential factors driving turbulence and understand their interactions.

2. Data and Methodology

In this study we have used the Stratosphere-Troposphere (S-T) wind profiler located at Kochi (10.04°N, 76.33°E), with a vertical resolution of 180 m for a period six years from March 20217 to December 2022 is used. Turbulent Kinetic Energy dissipation rate (ε) estimated using the spectral width method is used as the predictand and Level, Time of Day (TOD), Day of Year (DOY), horizontal wind Speed (V), vertical wind speed (w) and wind direction as predictors. Data points missing any of the six predictors were excluded, resulting in 611,070 data points. The method of estimation of ε is detailed in [1].

A cosine transformation is applied to the Julian Day and the 24-hour time to capture the cyclic nature of the day of the year and the time of day. This transformation ensures that the values of the day of the year (DOY) range between -1 (representing summer) and 1 (representing winter), and the values of the time of day (TOD) range between -1 (representing mid-noon) and 1 (representing midnight). The machine learning model was then trained using this data, with tuned hyperparameters. To evaluate the model's performance, we calculated the Pearson Correlation (R), Root Mean Square Error (RMSE), standard deviation, and bias between the modeled ε and the observed ε. Further analysis of the model was made using a XAI method which is a method to specifically explain tree based machine learning model like (Decision tree , Random Forest, XGBoost).

3. Results

The correlation obtained from the optimized model was 0.886 (Figure 1), with a bias of 5.4 x 10^{-7}, RMSE of 0.2919 and standard deviation of 0.4128. To study how each input variable affect the prediction in total we use the beeswarm plot (Figure 2). This plot demonstrates the impact of each feature of the model algorithm on the optimization of ε values, and the color bar indicates the ranking of global feature importance from high to low. It can be seen that the most important feature factor influencing the optimization of ε is the level. When the values of level are high, the SHAP value is low, indicating that the high value of the feature level contributes negatively to the modeled ε. While the high value of horizontal wind (V) is high, the feature encourages turbulence generation and the positive benefit contribution of the feature is greater than the negative benefit contribution. In the case of the DOY values a lower DOY (-1) is the summer season, pre-monsoon and monsoon season in our region, it contributes positively to the ε values while the winter season (1) contributes negatively which aligns well with the results from the work in [2]. In the case of vertical wind, a decrease in vertical wind speed values were found to be associated with stronger turbulence and an increase in vertical wind speed contributed to weaker turbulence.

4. Figures and Tables

4.1 Figure

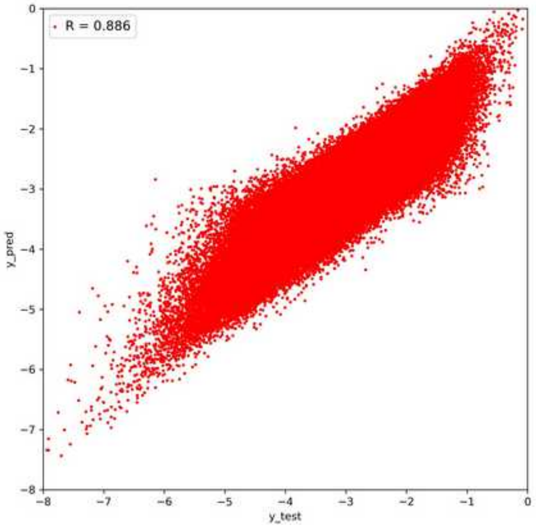

Figure 1 Correlation between modeled and observed data.

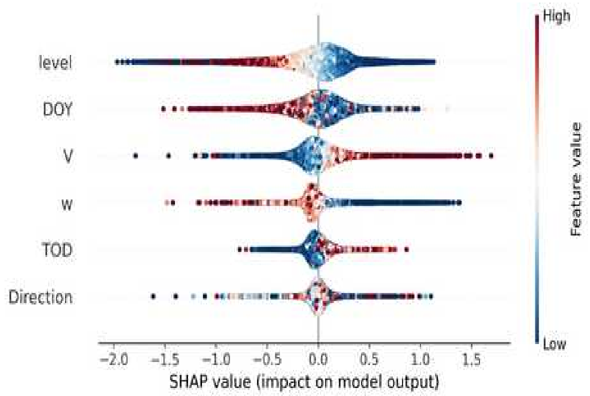

Figure 2 Beeswarm plot showing the contribution of each feature towards the estimation of ε.

5. Conclusions

In this work a machine learning approach was employed to better understand ε estimated from the S-T in relation to background conditions. XGBoost is the model used and we took ε as the predictand, and level, day of the year, time of day, horizontal wind speed, vertical wind, and wind direction as the predictors. A correlation coefficient of 0.886 between the modeled and training datasets after hyper tuning the model. After doing the SHAP analysis level was found to be the most influential variable. Regarding background wind conditions, horizontal wind speed emerged as the primary factor contributing to turbulence, with higher wind speeds correlating with increased turbulence. Conversely, lower vertical wind speed values were associated with higher turbulence

6. Acknowledgment

The S-T Radar at the Advanced Centre for Atmospheric Radar Research (ACARR), CUSAT, was established with funding from the Science and Engineering Research Board (SERB) under the Department of Science and Technology (DST), Government of India. We express our gratitude to the Ministry of Earth Sciences (MoES), Government of India, for providing financial support to maintain the S-T Radar Facility. The first author extends sincere thanks to CUSAT and DST for their fellowship support.

7. References

1. K. K. Ahana, K. Satheesan, S. Athulya, and A. Kottayil, “Turbulent kinetic energy dissipation rate in the lower troposphere using the 205 MHz radar at Kochi, India,” J. Atmos. Solar-Terrestrial Phys., vol. 251, Oct. 2023, Art. no. 106133. doi: 10.1016/j.jastp.2023.106133.

2. K. K. Ahana, Satheesan, and A. Kottayil, “Climatology of turbulent kinetic energy dissipation rate in the middle troposphere using the 205 MHz S-T radar at Kochi, India,” J. Atmos. Solar-Terr. Phys., vol. 256, Jan. 2024, Art. no. 106206.