NPS-PH-23-003

NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

PREDICTING OPTICAL TURBULENCE USING MACHINE LEARNING METHODOLOGY

by

Joseph Blau, Amanda Coleman, and Marthen Tamus

October 2023

Approved for public release. Distribution is unlimited.

Prepared for: Office of Naval Research

This research is supported by funding from the Naval Postgraduate School, Naval Research Program (PE 0605853N/2098). NRP Project ID: NPS-23-N106-A

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED	
21 Oct 2023	Technical Report	START DATE 24 Oct 2022	END DATE 21 Oct 2023

4. TITLE AND SUBTITLE
Predicting Optical Turbulence using Machine Learning Methodology

5a. CONTRACT NUMBER	5b. GRANT NUMBER	5c. PROGRAM ELEMENT NUMBER
		0605853N/2098
5d. PROJECT NUMBER NPS-23-N106-A; W2324	**5e. TASK NUMBER**	**5f. WORK UNIT NUMBER**

6. AUTHOR(S)
Blau, Joseph, A.; Coleman, Amanda, R.; Tamus, Marthen

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Postgraduate School
1 University Circle
Monterey, CA 93943-5000

8. PERFORMING ORGANIZATION REPORT NUMBER
NPS-PH-23-003

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)	11. SPONSOR/MONITOR'S REPORT NUMBER(S)
Naval Postgraduate School, Naval Research Program; Office of Naval Research	NRP; ONR	NPS-23-N106-A

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release. Distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Measuring and predicting optical turbulence is difficult and requires specialized equipment. The NPS Meteorology Department has previously developed a model (NAVSLaM) to predict optical turbulence in the surface layer (up to ~100 m above the ocean or land) based upon atmospheric measurements using simple, robust sensors. On the other hand, the Physics Department has developed machine learning models of optical turbulence using atmospheric measurements. This research involves measurements of optical turbulence over many months using sonic anemometers that served as the baseline to compare prediction from the models. Atmospheric parameters such as air temperature, wind speed, humidity at two different heights as well as solar flux and ground temperature were simultaneously collected. Those data were used as inputs for NAVSLaM and the machine learning models to predict optical turbulence. We then compared the performance of these prediction models to each other by calculating the root-mean-square error with respect to the baseline data from the sonic anemometers. The results from this research will help determine which model is more reliable for the given environment. Overall, the ML model appeared to work better than NAVSLaM for predicting the optical turbulence values that we observed. However, NAVSLaM is a more general model that should work well in a variety of environments. An accurate machine learning model of optical turbulence could significantly improve forecasts of directed energy weapon effectiveness. Eventually, it could even be used in an operational scenario to make real-time predictions of turbulence and its impact on directed energy weapon performance.

15. SUBJECT TERMS
High energy lasers (HELs), atmospheric propagation, optical turbulence, machine learning (ML)

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES
a. REPORT UNCLASSIFIED	b. ABSTRACT UNCLASSIFIED	C. THIS PAGE UNCLASSIFIED	UU	38

19a. NAME OF RESPONSIBLE PERSON	19b. PHONE NUMBER *(Include area code)*
Joseph Blau	831-656-2332

STANDARD FORM 298 (REV. 5/2020)
Prescribed by ANSI Std. Z39.18

THIS PAGE INTENTIONALLY LEFT BLANK

NAVAL POSTGRADUATE SCHOOL
Monterey, California 93943-5000

Ann E. Rondeau
President

Scott Gartner
Provost

The report entitled "Predicting Optical Turbulence using Machine Learning Methodology" was prepared for the Office of Naval Research and funded by the Naval Research Program.

Further distribution of all or part of this report is authorized.

This report was prepared by:

BLAU.JOSEPH.A.1228749016 Digitally signed by BLAU.JOSEPH.A.1228749016 Date: 2023.11.01 13:36:02 -07'00'

Joseph Blau
Research Associate Professor

COLEMAN.AMANDA.ROSE.1511795628 Digitally signed by COLEMAN.AMANDA.ROSE.1511795628 Date: 2023.11.01 13:51:25 -07'00'

Amanda Coleman
Research Associate

Reviewed by:

NARDUCCI.FRANCESCO.A.1229136050 Digitally signed by NARDUCCI.FRANCESCO.A.1229136050 Date: 2023.11.01 16:36:24 -07'00'

Frank Narducci, Chairman
Physics

Released by:

SMITH.KEVIN.B.1230026366 Digitally signed by SMITH.KEVIN.B.1230026366 Date: 2023.11.03 06:06:17 -07'00'

Kevin B. Smith
Vice Provost for Research

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Measuring and predicting optical turbulence is difficult and requires specialized equipment. The NPS Meteorology Department has previously developed a model (NAVSLaM) to predict optical turbulence in the surface layer (up to ~100 m above the ocean or land) based upon atmospheric measurements using simple, robust sensors. On the other hand, the Physics Department has developed machine learning models of optical turbulence using atmospheric measurements. This research involves measurements of optical turbulence over many months using sonic anemometers that served as the baseline to compare prediction from the models. Atmospheric parameters such as air temperature, wind speed, humidity at two different heights as well as solar flux and ground temperature were simultaneously collected. Those data were used as inputs for NAVSLaM and the machine learning models to predict optical turbulence. We then compared the performance of these prediction models to each other by calculating the root-mean-square error with respect to the baseline data from the sonic anemometers. The results from this research will help determine which model is more reliable for the given environment. Overall, the ML model appeared to work better than NAVSLaM for predicting the optical turbulence values that we observed. However, NAVSLaM is a more general model that should work well in a variety of environments. An accurate machine learning model of optical turbulence could significantly improve forecasts of directed energy weapon effectiveness. Eventually, it could even be used in an operational scenario to make real-time predictions of turbulence and its impact on directed energy weapon performance.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

This chapter is adapted from M. Tamus, "Comparison of Optical Turbulence Prediction Models," M.S. Thesis, Naval Postgraduate School, Monterey, CA, USA, 2022 [1].

A. BACKGROUND

A laser produces a coherent, monochromatic, and very narrow beam of light by stimulated emission of radiation. High energy lasers (HELs) can precisely deliver lethal energy over ranges on the order of several kilometers [1]. Additionally, the beam spot size at the target is on the order of a few centimeters across, minimizing the prospects of collateral damage. An HEL can accurately engage highly maneuverable targets. It is also capable of causing a specific, predetermined amount of damage to the target because the output power of the laser is controllable. The enormous magazine capacity and low cost per engagement are other advantages for HEL weapons. However, compared with kinetic weapons, HELs are more strongly affected by the atmosphere, in particular optical turbulence, as discussed in the following section.

Optical turbulence has various effects on laser beam propagation, including beam breakup, wander and scintillation. Therefore, estimating the level of turbulence along the beam path is essential to predict HEL performance and to determine appropriate turbulence mitigation measures [2]. However, measuring or predicting turbulence levels is not easy apart from the fact that the strength of the turbulence along the beam path can vary spatially and temporally. Also, under certain environmental conditions, such as at sea, the delicate equipment required to measure point turbulence often does not work correctly. Under these environmental conditions, estimating optical turbulence from standard meteorological parameters that can be measured using more robust equipment or forecasted by a numerical weather prediction model is a more feasible alternative [3].

One turbulence model is the Navy Atmospheric Vertical Surface Layer Model (NAVSLaM) which is developed by the Department of Meteorology of the Naval Postgraduate School (NPS). NAVSLaM estimates turbulence levels in the surface layer, up to ~100 m above the ocean or land, based upon meteorological parameters such as temperature, wind speed, and humidity. This model can predict turbulence well in a

homogeneous environment but may not work as well in nonhomogeneous environments (e.g., the maritime coastal environment) [4].

Previous efforts by the Directed Energy Group at NPS have shown that machine learning (ML) models can predict turbulence even in a nonhomogeneous environment [5]. This effort measured atmospheric parameters at one level using more robust sensors from the roof of Spannagel Hall at NPS over many months. The collected data were then used to select and train an ML model. After selecting the best performing model, a bagged ensemble of regression trees, the data were then used to predict the refractive index structure parameter (C_n^2), a critical parameter for quantifying the strength of atmospheric turbulence. Checking the performance of our ML regression model using a test set showed promising predictive performance. However, this method still needs to be validated using data at two heights as well as from varying locations.

The goal of this project is to collect atmospheric data from two different heights in a coastal environment to develop an ML model that can accurately predict turbulence. Furthermore, we want to directly compare the predictive performance of the ML model to NAVSLaM in nonhomogeneous environments. We conducted an experiment with the help of our thesis student, Marthen Tamus, to measure point turbulence at two levels as well as to collect meteorological data such as wind speed, sonic temperature, air temperature, soil temperature, barometric pressure, and solar flux [15]. Though still collecting data, this project uses three months of processed data collected at the Naval Research Laboratory Coastal Environmental Observation Station (NRL-CEOBS), located on the edge of Monterey Bay Beach near Watsonville. Using these simple atmospheric measurements and point turbulence measurement data, we developed an ML model to predict optical turbulence, and compared the results to NAVSLaM predictions.

II. ATMOSPHERIC OPTICAL TURBULENCE

This chapter is adapted from M. Tamus, "Comparison of Optical Turbulence Prediction Models," M.S. Thesis, Naval Postgraduate School, Monterey, CA, USA, 2022 [1].

Optical turbulence is characterized by random fluctuations in the temperature of small packets of air called "eddies." These fluctuations are due to solar heating and wind shear. The temperature fluctuations correspond to fluctuations in the density of the eddies, and thus their refractive index. The fluctuations in the refractive index affect the propagation of optical waves, including laser beams, through the atmosphere [7].

The intensity of optical turbulence is quantified by the index of refraction structure constant, C_n^2. [8] The behavior of C_n^2 at a point along the propagation path can be deduced from the temperature fluctuations at that point due to the strong correlation of temperature to refractive index. This is important since temperature fluctuations are usually much easier to measure than direct refractive index measurements. Then, C_n^2 can be expressed in terms of the temperature structure constant C_T^2 [8]:

$$C_n^2 \approx \left(79 \times 10^{-6} \frac{P}{T^2}\right)^2 C_T^2, \tag{1}$$

where P is the pressure (mbar) and T is the temperature (K). C_T^2 may be obtained from point measurements of temperature fluctuations obtained from sonic anemometers.

A. MEASUREMENTS OF TURBULENCE USING SONIC ANEMOMETERS

As previously discussed, turbulent eddies have different temperatures and different densities that lead to fluctuations in the index of refraction. One way to measure the structure parameter of the index of refraction is by using a sonic anemometer, a sensor that can sample rapidly (e.g., 20 Hz) the temperature fluctuations of eddies that are passing between its sonic transducers . A sonic anemometer uses three pairs of non-orthogonally arranged transducers to transmit and receive ultrasonic waves in opposite directions. The speed of sound can then be calculated by measuring the flight time of the sonic signal between a pair of transducers, and from that we can estimate the sonic temperature, which will rapidly vary as turbulent eddies pass through the sensor.

Figure 1 shows an example of sonic temperature fluctuations taken by a sonic anemometer versus time.

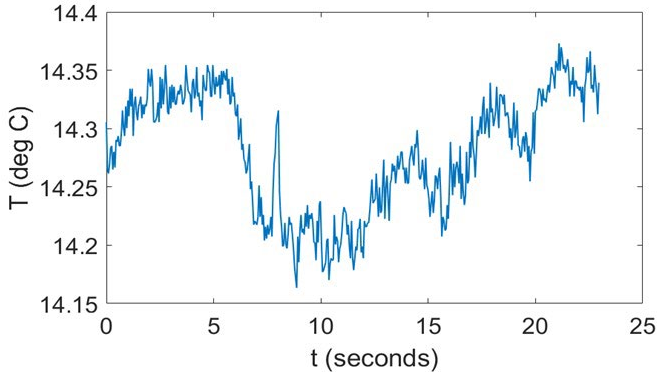


Figure 1. Sonic temperature fluctuations measured by a sonic anemometer versus time.

The amplitudes and frequencies of the temperature fluctuations indicate the size of the turbulent eddies passing through the sensor. Larger eddies produce larger amplitudes and lower frequency fluctuations compared to smaller eddies. If we take the power spectral density (PSD) of temperature as a function of time and plot it on a log-log plot as shown in Figure 2, then the frequency range where the slope is -5/3 defines the inertial subrange according to Kolmogorov's turbulence theory [8].

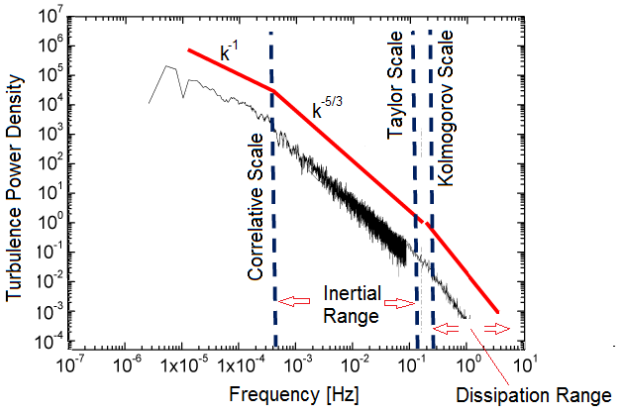


Figure 2. Power Spectral Density versus frequency. Source: [9].

More specifically, within the inertial subrange, the amplitude of the power spectral density PSD should be related to the frequency *f* according to:

$$PSD\ (f) = C_T^2\ f^{-5/3} \tag{2}$$

Graphing this on a log-log plot as in Figure 2, we should get -5/3 slope [4]. The value of C_T^2 is then obtained by the fitting line to the PSD. From equation (1) we then can estimate the value of C_n^2.

B. MODELING OF TURBULENCE USING NAVSLAM

NAVSLaM is an optical turbulence prediction model developed by the NPS Meteorological Department and is based on Monin-Obukhov similarity theory (MOST) [10], which assumes that the atmosphere layers near the surface are horizontally homogeneous and stationary. The original version of NAVSLaM only requires atmospheric data at a single level and is valid for over-the-ocean surface application. This version can produce profiles of the refractive index structure parameters, C_n^2, at any height above the ocean surface up to 100 meters.

A newer version of NAVSLaM that requires inputs of atmosphere data at two different heights within the atmospheric boundary layer was recently developed and can estimate valid C_n^2 profiles not only over the ocean but also over the land [11]. The input atmospheric data such as wind speed, air temperature and humidity can be measured or modeled. In this project, we provided this version of NAVSLaM with data measured by two sets of sensors at different levels. We then compared the prediction of the two-level NAVSLaM with a machine learning model trained using the same data [15].

III. EXPERIMENTAL SETUP

This chapter is reproduced from M. Tamus, "Comparison of Optical Turbulence Prediction Models," M.S. Thesis, Naval Postgraduate School, Monterey, CA, USA, 2022 [1].

This experiment aimed to collect atmospheric parameters (air temperature, ground temperature, solar flux, wind speed, and humidity), estimate the point refractive index structure parameter C_n^2 (from sonic anemometer data), and use them as predictor and response inputs, respectively, to train the machine-learning optical turbulence prediction models. The predicted values of C_n^2 are then compared with the point-estimate values and the predictions from NAVSLaM using the same measured atmospheric parameters. For the purposes of that comparison, an experiment was designed and set up on the Naval Research Laboratory Coastal Environmental Observation Station (NRL-CEOBS), located at the Monterey Bay Academy, near Watsonville, CA (Figure 3).

Figure 3. Experimental Platform at the Naval Research Laboratory Coastal Environmental Observation Station (NRL-CEOBS).

The setup consists of two sets of sonic anemometers and hygrometers at two different levels (the height of the lower sonic anemometer is 1.35 meters, and the upper sonic anemometer is 4.7 meters above the ground), an infrared radiometer to measure ground temperature, a solar flux monitor, a GPS receiver, and a data logger. The whole system was mounted on a tripod with both sonic anemometers facing Monterey Bay in the direction of typical onshore winds from the bay. Data from the sensors were recorded with a data logger onto a Secure Digital (SD) card and were collected continuously from August 26 to the end of November 2022.

A. CSAT3B THREE- DIMENSIONAL SONIC ANEMOMETER

The CSAT3B uses ultrasonic sound waves to determine speed of sound and instantaneous wind speed by measuring the travel time of sound waves between a pair of transducers through the air. The sonic temperature then can be derived from the speed of sound. The ability to take measurements with very rapid frequency, 20 Hz or higher, makes sonic anemometry a popular choice for the study of optical turbulence.

We used the sonic temperature recorded by two CSAT3Bs at different heights (at 1.35 m and 4.7 m above the ground) to derive point estimates of the refractive index structure parameter, C_n^2, at the respective heights of the anemometers. To convert the sonic temperature data to C_n^2, first, we organized the data into folders based on the collecting date. We then divided the data into tranches with 2048 samples each. Next, we applied Welch's method to each tranche to obtain the power spectral density (PSD). The relationship between the frequency-dependent temperature structure function C_T^2 and PSD amplitude S_T is given by:

$$C_T^2(f) = 4\left(\frac{2\pi}{U}\right)^{\frac{2}{3}} S_T(f) f^{5/3}, \tag{3}$$

where U is the mean wind speed as measured by the sonic anemometers for a particular tranche and f is the temporal frequency of the corresponding power spectrum. An average value of C_T^2 across the spectrum was obtained by integrating both sides of this equation with respect to frequency:

$$\langle C_T^2 \rangle = \left(\frac{4}{f_2 - f_1}\right)\left(\frac{2\pi}{U}\right)^{\frac{2}{3}} \int_{f_1}^{f_2} S_T(f) f^{5/3} df \quad (4)$$

The upper and lower limits of integration were chosen to be 5 Hz and 1 Hz, respectively; we found (empirically) that these limits are usually inside of the inertial subrange for a wide range of atmospheric conditions. The reported values of C_n^2 of were derived from $\langle C_T^2 \rangle$.

B. OTHER SENSORS

We used a HygroVUE 5 sensor to measure air temperature and relative humidity at two different heights. Measurement of the air temperature and relative humidity served as input data both for the machine learning model and for NAVSLaM. We used a SN500SS net radiometer to measure solar radiation flux, and we used SI-111 infrared radiometer to measure the ground temperature. A GPS receiver was used to record the local time, and all the collected data was recorded on a CR6 data logger.

For each day the collected files were merged and transformed into a MATLAB data file. Then this data file was used to calculate the experimental values of C_n^2**.** The data in the MATLAB file and the experimental values of C_n^2 from the sonics were then used as the inputs and response, respectively, for training the machine learning model.

IV. MACHINE LEARNING

This chapter is adapted from M. Tamus, "Comparison of Optical Turbulence Prediction Models," M.S. Thesis, Naval Postgraduate School, Monterey, CA, USA, 2022 [1].

A. INTRODUCTION

Physics-based models used to characterize turbulence in maritime or coastal environments already exist [4], [11]. However, physical models that use Monin-Obukubov similarity theory (MOST) assume horizontally homogeneous environmental conditions, Because of this assumption, modeling turbulence in more complex environments is a challenge. For example, turbulence above the deck of a ship is not only affected by the sea surface temperature and the air temperature above it, but also by the heating of the ship's deck and the airflow around the superstructure.

In such an environment, an alternative is to use an ML regression model to predict turbulence levels based on various meteorological parameters such as air temperature, wind speed, and humidity. Unlike a physical model based on MOST, an ML model uses predictors (i.e., the input meteorological parameters) to determine the response (i.e., the output turbulence level) by analyzing a large training data set to develop relationships between the predictors and the response.

B. OVERVIEW OF MACHINE LEARNING REGRESSION ANALYSIS

Machine Learning (ML) is an incredibly broad field within Artificial Intelligence that utilizes statistical methodology to learn from data. Machine learning can be classified as supervised learning and unsupervised learning, depending on the application. Regression analysis is a supervised ML technique used to estimate the relationship between predictor and continuous response variables. Linear regression models are used when there is a linear relationship between predictors and the response. Despite their simplicity and accessibility, linear regression models are often inadequate when the response variable cannot be expressed as a linear combination of predictors. In many such cases, ML methods that do not rely on linear relationships can be used.

To create a machine learning model, the data are partitioned into a training and test set. For our model, we randomly selected 70% of our data for training and will use the

remaining 30% as a test set to check model performance. Similarly, k-fold cross validation was used to improve our model's performance. Selecting a k value between 5 and 10 has been shown to help navigate the bias-variance trade off discussed later in this section. For our large dataset, we chose to use k=5. Increasing the number of folds improved the performance slightly, but drastically increased runtime.

C. MODEL SELECTION

A standard metric used to quantify model performance is the mean square error (MSE) and the root mean square error (RMSE), which evaluates the difference between observed and predicted values for a given data set. Several ML algorithms were explored, including linear regression, regression ensembles, Gaussian process models, support vector machines, and simple neural networks. Several were found to have comparatively low root mean squared error (RMSE), but the bagged ensemble of regression trees performed the best.

1. Decision Tree

A decision tree is a type of supervised machine learning model frequently used for classification and regression. A decision tree can be represented by a flowchart-like graph consisting of nodes connected by branches (Figure 4).

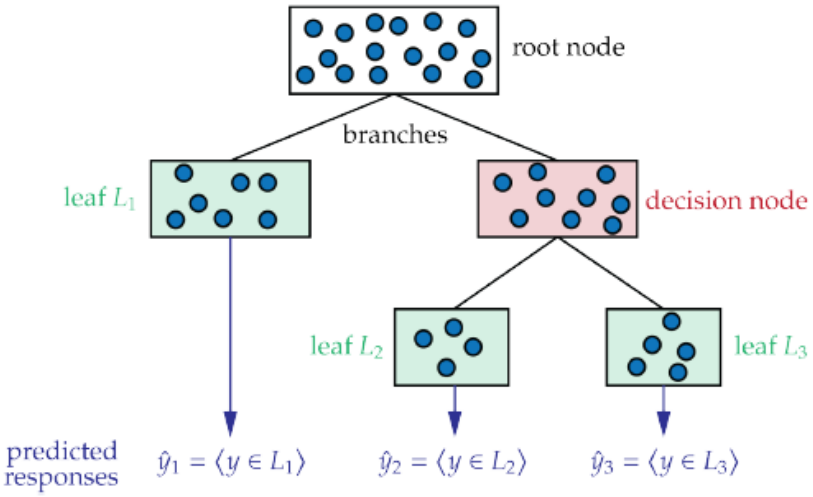


Figure 4. Decision Tree Example. Source: [5].

Because of their simplicity, decision trees are often referred to as weak learners - models that will perform slightly better than randomly guessing. A decision tree is created by first assigning a set of training data to an initial node, called the root node. The data is then split into subsets as described below. Each node beneath the root node contains a subset of the original data. A node where splitting occurs is referred to as a decision node. Terminal nodes at the end of each branch are called leaf nodes. The resulting decision tree can then be applied to new data.

2. Bagged Ensemble of Regression Trees

Bootstrap Aggregation, commonly referred to as bagging, is a resampling technique used to improve predictive performance of weak learners. A bagged ensemble can be constructed by combining trained, weak learners. For example, as seen in Figure 5, the training data, *D*, are split into *k* subsets with replacement. These *k* subsets are then used to create *k* different models. Finally, the algorithm aggregates the *k* weak learners into a final model. Many trained decision trees are combined to produce a bagged ensemble of regression trees. Creating a tree ensemble greatly reduces variance while maintaining a consistent bias.

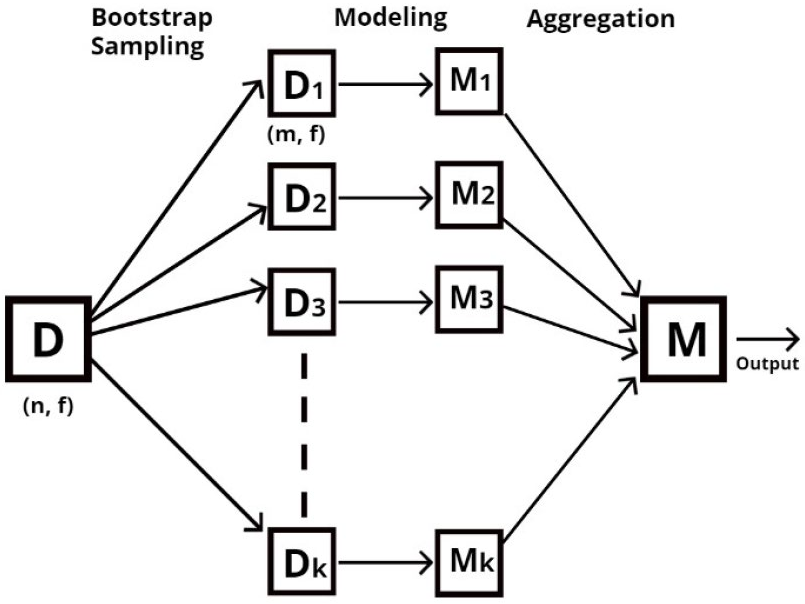


Figure 5. Ensemble of Bagged Trees Model Diagram. Source: [21].

IV. RESULTS AND ANALYSIS

This chapter is reproduced from M. Tamus, "Comparison of Optical Turbulence Prediction Models," M.S. Thesis, Naval Postgraduate School, Monterey, CA, USA, 2022 [1].

A. DATA PREPROCESSING

After collecting data for approximately three months, we prepared the data before training machine learning models. First, we derived the structure parameter of the index of refraction C_n^2 from sonic anemometer measurements based on the Kolmogorov theory. To do this, we took the power spectral density (PSD) of the sonic temperature over 102 second intervals (using 2048 observations acquired at 20 Hz) to determine the average of C_T^2 (Equation 15). We then derived C_n^2 with the same interval for the entire measurement data using Equation 6. We obtained 842 observations of C_n^2 for each day or approximately 70,000 observations over the entire three-month experiment from each sonic anemometer [15].

We had several challenges when we processed the data. The first challenge is that, when we calculate C_n^2 from the measurement of the upper and lower sonic anemometers, the data sampling is reduced, and the time interval is not the same as the measured sonic temperature. The time interval also differs between the sonic anemometers and the other sensors because we sampled the sonics at 20 Hz while the other sensors were sampled at 0.2 Hz. The second challenge is that there are some data points where the calculation of C_n^2 resulted in "not a number" (NaN) values, which is not a data type accepted by the MATLAB machine learning tool. These NaN values correspond to invalid data. The third challenge is that Kolmogorov theory assumes the data falls within the inertial subrange, where the slope of the PSD versus frequency is -5/3 on a log-log plot; not all our calculated PSDs have this expected -5/3 slope. To solve the first problem, we interpolated all the other measurement data to the estimated C_n^2 data and sorted by time stamp. To solve the second problem, we removed all the data rows that yielded NaN values for C_n^2. To solve the third problem, we filtered out all the data with PSD slopes less than -2 and greater than -1 .

Figure 6 shows an example plot of the predicted C_n^2, the measured sonic temperature, wind speed, and calculated PSD slope over 24 hours on October 30, 2022.

The PSD slope fluctuates around -5/3 as expected. The C_n^2 curve also follows the expected diurnal trend, with the values rising after dawn, reaching a peak at noon, and dropping again in the afternoon.

Finally, we compiled all the interpolated measurement data into a single file that can be used to train the MATLAB machine learning model. For our model, we used the logarithm of C_n^2 as a response parameter, because C_n^2 can vary over several orders of magnitude.

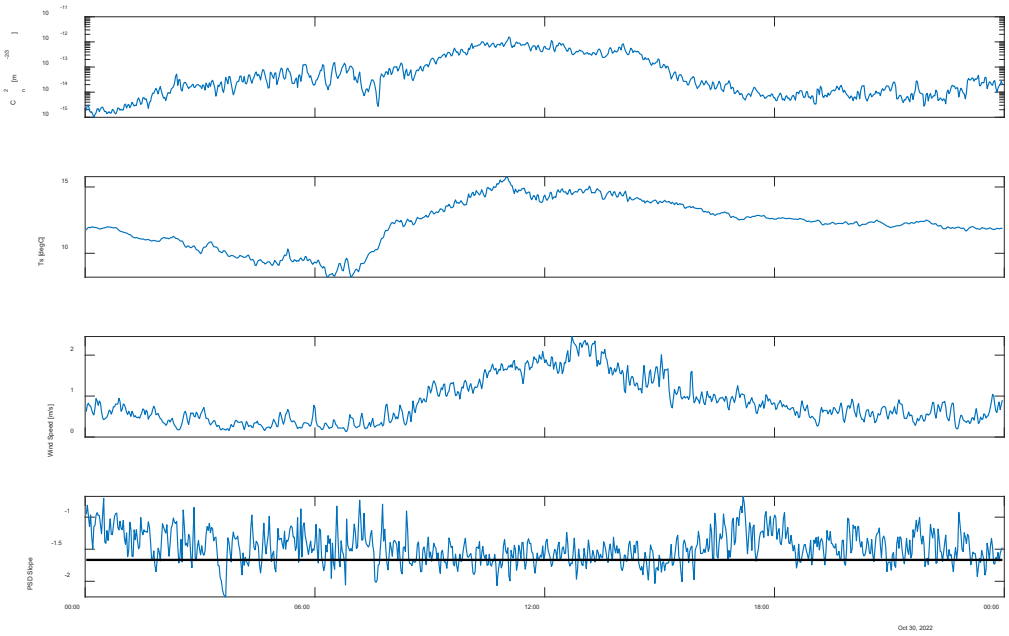


Figure 6. C_n^2 derived from Sonic Anemometer data plotted along with the corresponding sonic temperature, wind speed and PSD slope of the sonic temperature.

B. MODEL SELECTION AND OPTIMIZATION

We first imported the preprocessed data into MATLAB. Since the training process takes a lot of time, we selected only 50% of the total data collected from 26 August to 3 November 2022 to initially run with all the regression methods. Furthermore, to protect against overfitting, we selected the five-fold cross-validation. We then ran the application to train all 27 regression methods in parallel. The application automatically reported the

RMSE of each model when the training process was complete; MATLAB highlighted the method with the lowest RMSE, indicating the best-performing model using this metric.

Figure 7 shows the training results for many ML models for 50% of the total data. The model panel on the left shows the sorted models based on the RMSE score after the training. Three models with the lowest RMSE score were Wide Neural Network, Ensemble Bagged Trees, and Exponential GPR. Figure 7 also displays on the right the response plot (C_n^2) for the observed values (blue points) and predicted values (orange points) for the best model (Wide Neural Network).

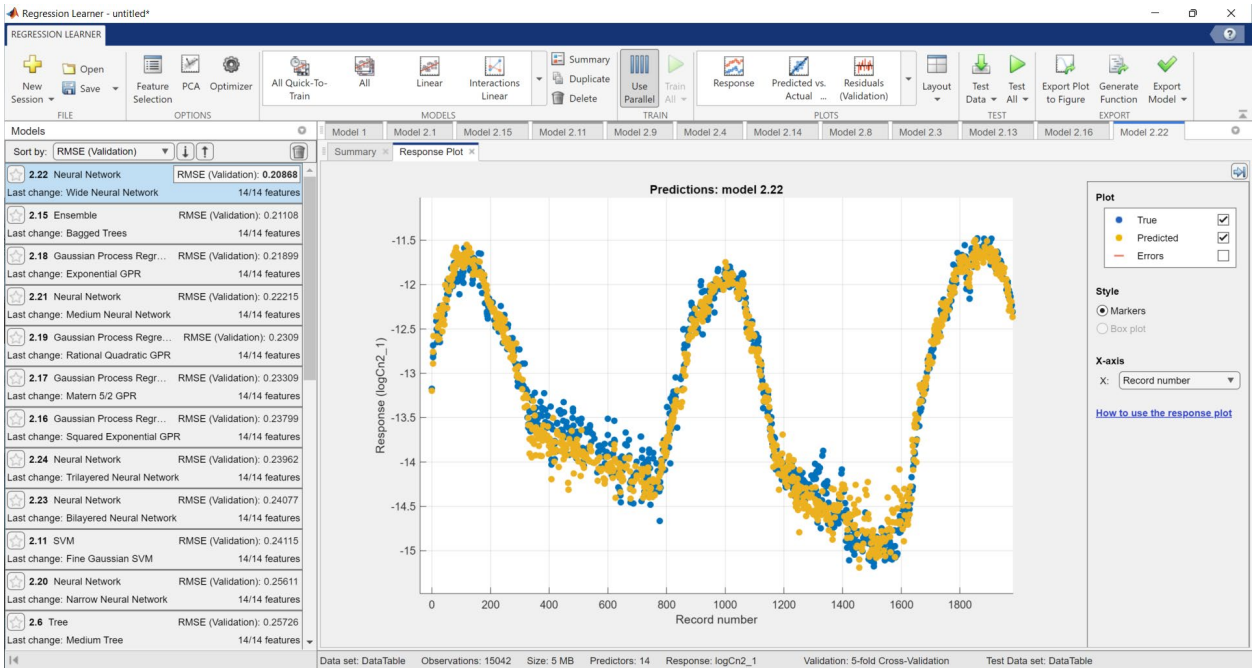
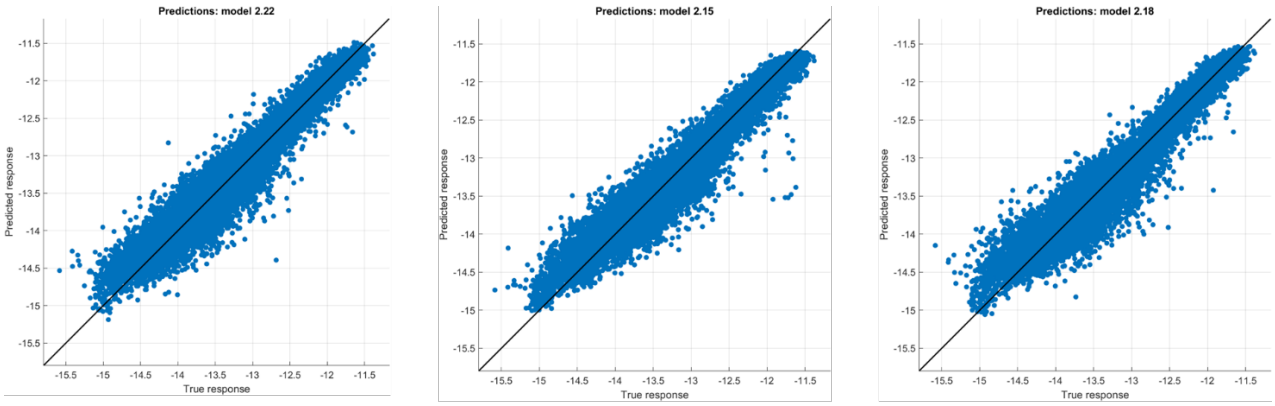


Figure 7. Training result using 50 % of randomly selected data set.

In addition, we generated several plots that can be used to observe the model's performance. Figure 8 plots the predicted response versus the observed response for the same data set for the three models; in other words, if the model agreed perfectly with the measurement, then all the points would be along this line. The three figures look similar, showing that the models tend to underestimate and overestimate the large C_n^2 values and overestimate the smaller C_n^2 values. However, this is the expected behavior if the extremes in the measured response data are caused by noise.



a) b) c)

Figure 8. Predicted vs. Measured response of log C_n^2(Date: November 13, 2022). a) Wide Neural Network regression method with RMSE=0.209; b) Ensemble bagged trees regression method with RMSE=0.211; c) Gaussian process regression method: Exponential GPR with RMSE=0.219

Since the three models have very similar RMSE values, and the wide neural network regression model and the exponential GPR model required a long time for model training compared to the ensemble model, we decided to use the ensemble bagged trees model which has a slightly larger value of RMSE but runs much faster. Moreover, a previous experiment has proven by a series of trial with a larger data set that the bagged ensemble of regression trees has a good performance in predicting C_n^2 [5]. Therefore, we decided to train our data set with the ensemble bagged trees model.

C. MACHINE LEARNING PREDICTION RESULTS

To conduct ML model training and to perform validation tests over the collected meteorological data set, we randomly select 70% of the existing data as training data and the remaining 30% as test data. The training data was used to train the ensemble bagged trees ML model using the following as predictors: air temperature, humidity, and wind speed (each measured at two different altitude levels), the solar flux, the ground temperature, and the altitude above the ground; the response was the logarithm of C_n^2 measured by two sonic anemometers at different heights. For a given set of predictors, the altitude above the ground was set to the height of the sonic anemometer that was used to compute C_n^2 [15].

Figure 9 shows the prediction of C_n^2 by the trained model using 30% of the test data and the corresponding point C_n^2 measured by the sonics from September 30 to November 30, 2022. To be clear, the test data was not used in the training process. The $\log(C_n^2)$ values of the upper sonic anemometer are represented in the first plot by the red line, overlaying the $\log(C_n^2)$ values from the ML model in blue. In the middle plot, the $\log(C_n^2)$ values of the lower sonic anemometer and ML model are also represented in red and blue, respectively. The green line in the bottom plot represents the temperature difference measured by the upper and lower hygrometers. This temperature difference is important to include when comparing the ML model to NAVSLaM (in the next section) and is included here for that reason.

The upper two plots of Figure 9 show that the ML model predicted similar $\log(C_n^2)$ values to the corresponding values from the sonic anemometers for the entire observation period. The RMSE for predicted $\log(C_n^2)$ from the upper test data is 0.209 and from the lower test data is 0.194. Moreover, as we can see in Figure 9(b), which is a magnification of the observations between 23 and 28 October 2022, that both predictions also follow a typical diurnal pattern, with peaks of C_n^2 values occurring during warmer afternoons and lower C_n^2 values observed between cooler evenings and early mornings. We also noticed that the C_n^2 values for the lower of the two heights are usually greater than the upper height. This follows a well-known trend that turbulence is normally greater closer to the ground [2], [7].

The temperature differences (upper hygrometer air temperature minus lower hygrometer air temperature) in the lower plot generally showed larger negative values (up to -2 degrees Celsius) during the day, as the upper sensor is further away from the warm ground. At night, the air temperature difference was near neutral for the first half of the observation period; during the second half, the nighttime temperature at the location of the upper hygrometer tended to be slightly higher than the lower hygrometer. The positive temperature differences are more than +2 degrees Celsius on several days during this period. However, the ML model predictions agreed well with the measured values of C_n^2 and are independent of the temperature differences.

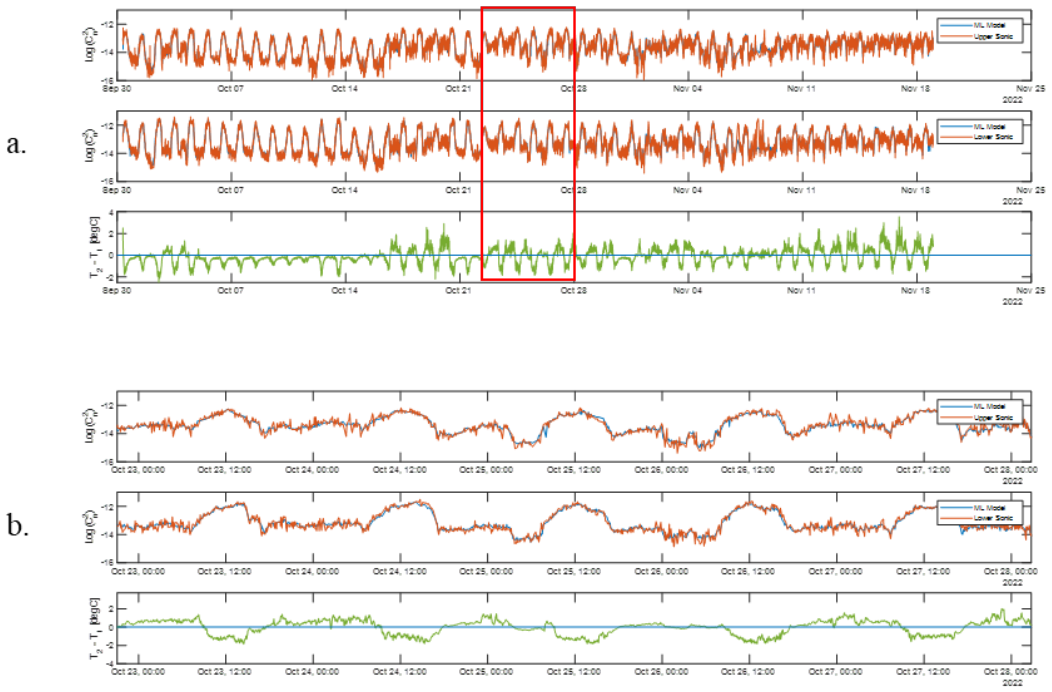


Figure 9. a. The measured $\log(C_n^2)$ values overlaying the predicted $\log(C_n^2)$ values by the ML model through the entire period of observations. b. The same plot is zoomed in on the period of 23 to 28 October 2022.

Figure 10 shows the predicted versus measured values of $\log(C_n^2)$ for the same data at the two different heights; the left plot is for the upper observation point, and the right is for the lower observation point. The color of each data point corresponds to a temperature difference between the upper and lower observation points, whose value is indicated by the color bar on the right side of each plot. As seen, the $\log(C_n^2)$ values in the upper observation point are distributed at slightly smaller values than those in the lower observation point. The highest value at the upper observation point is not more than -12, while at the lower observation point, it is greater than -12. The lowest $\log(C_n^2)$ value observed at the upper observation point is less than -15, while the $\log(C_n^2)$ value at the lower observation point is not less than -15.

Also shown on the plots of this figure is a red line that represents perfect agreement; in other words, all the points would lie along this red line if there were perfect agreement between ML model predictions and the observed values. Overall, the predicted values for upper and lower C_n^2 have good agreement with the measured C_n^2 values as the observations

cluster near the red diagonal line. The ML model only slightly overestimates the small values of $\log(C_n^2)$, which is similar to the behavior found in [5]. Otherwise, the scattering of the points is evenly distributed across the line of perfect agreement. This indicates that the ML model is largely unbiased.

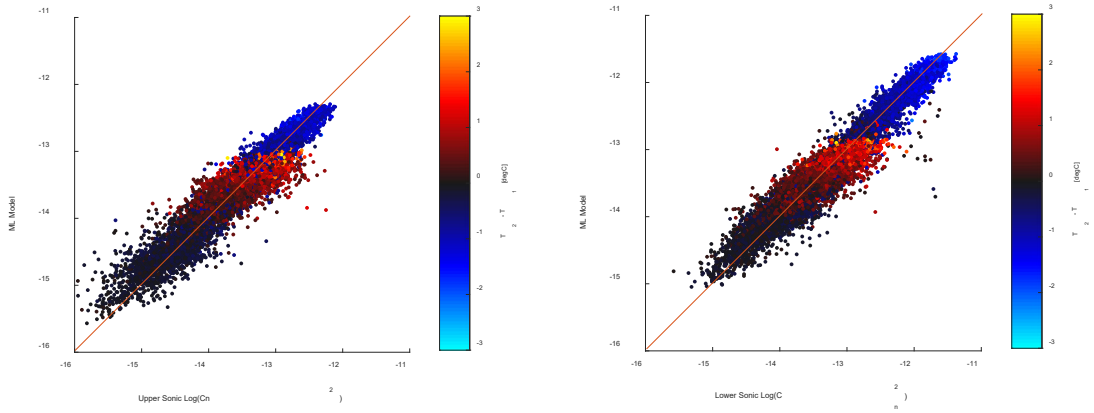


Figure 10. The $\log(C_n^2)$ values predicted by the ML model versus the values measured by the upper and lower sonic anemometers.

D. COMPARISON WITH NAVSLAM

For comparison purposes, we used the same meteorological data (except for solar flux and ground temperature) for training the ML model to estimate the value of C_n^2 by the NAVSLaM model [15]. We plotted the $\log(C_n^2)$ values estimated by NAVSLaM in the same graphs with the $\log(C_n^2)$ values measured by the upper and lower sonic anemometer, respectively. Figure 11 shows the resulting plots, which are analogous to the plots in Figure 9; the difference is the blue curves in the two upper plots now represent the C_n^2 of NAVSLaM estimates, replacing the C_n^2 values predicted by the ML model in Figure 9. Compared to the ML model predictions, NAVSLaM estimates are seen to not agree as well with the measured values of C_n^2, especially during periods of weaker turbulence as measured by the sonics. However, the NAVSLaM estimates at upper and lower observation levels still follow a typical diurnal pattern. The agreements are better when the temperature difference values result in unstable conditions (when the temperature difference is negative; i.e., when the temperature of the lower sonic is warmer than the upper sonic), as

shown in Figure 11(b), which zoomed in the results of the observations between 23 and 28 October 2022. Figure 11 also shows that the NAVSLaM model tends to poorly estimate the value of C_n^2 (relative to the sonic measured values) when the temperature differences are close to zero; on the other hand, it tends to underestimate the C_n^2 values when the temperature difference values are positive.

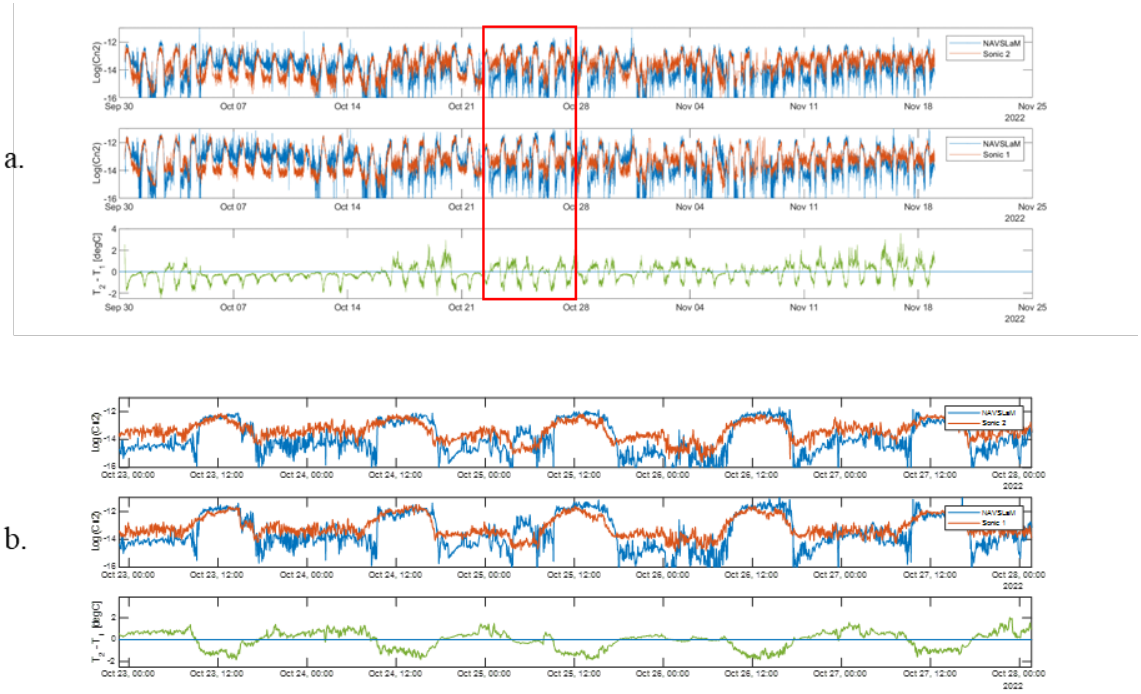


Figure 11. a. The measured $\log(C_n^2)$ values overlaying the estimated $\log(C_n^2)$ values by NAVSLaM model through the entire period of observations. b. The same plot is zoomed in on the period of 23 to 28 October 2022

Figure 12 shows the plots of NAVSLaM-derived $\log(C_n^2)$ versus the observed $\log(C_n^2)$ measured by sonic anemometers at two different observation levels; the left plot is for the upper observation level, and the right is for the lower observation level. The color of each data point corresponds to a temperature difference between the upper and lower observation points, whose value is indicated by the color bar on the right side of each plot. From Figure 12, we can see that the observations spread around the diagonal line that corresponds to perfect correlation. NAVSLaM tends to be less accurate in estimating the value of C_n^2 when the temperature difference is close to zero, as shown by the wide scattering black points. NAVSLaM tends to underestimate C_n^2 values when the temperature

differences are positive (reddish colors); the agreement seems to increase as the temperature difference increases in a positive direction, as shown by the dark red and near orange points under the diagonal line. NAVSLaM has a reasonable estimation of C_n^2 when the temperature differences are negative numbers; the agreement seems to increase as the difference increases in a negative direction, as shown by the clustering of blue points.

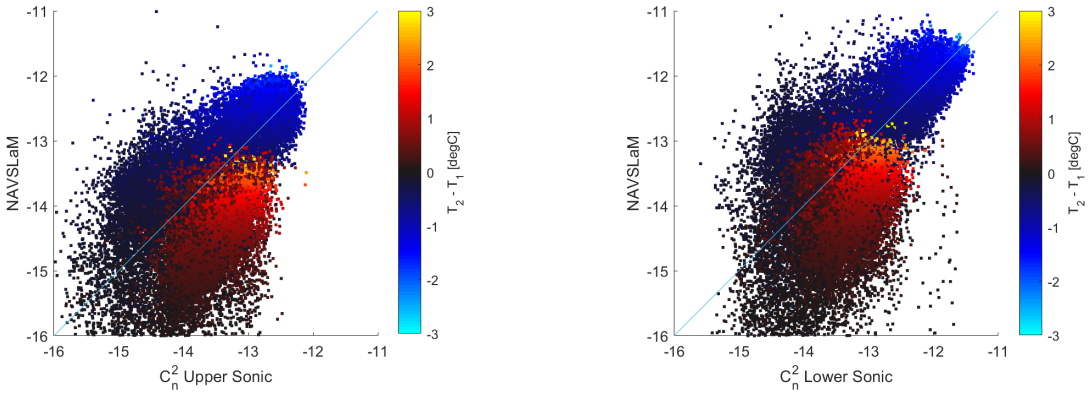


Figure 12. The log(C_n^2) values estimated by NAVSLaM versus the observed C_n^2 measured by sonic anemometers at two different heights.

A significant disagreement between the C_n^2 values predicted by NAVSLaM and the values measured by the CSAT3B were mainly observed when there was little or no difference in air temperature between the two levels. In these near-neutral conditions, the MOST theory that NAVSLaM is based upon is known to work not so well [10]. In addition, turbulence is expected to reach its lowest values in the neutral region, while the capability of CSAT3B to detect C_n^2 is not less than 10^{-16} $m^{-2/3}$ due to their noise floor. One of these two possibilities or a combination of the two is believed to be the reason for this significant disagreement.

To further explore the relationship between temperature differences and C_n^2 values, we plotted the logarithm of C_n^2 values measured by each sonic anemometer and C_n^2 values estimated by NAVSLaM versus temperature differences [15]. The top two plots in Figure 13 show the estimated value of C_n^2 by NAVSLaM versus the temperature difference measured at the upper and lower hygrometers. The bottom two plots show the C_n^2 values based on the CSAT3B measurements versus the mean temperature difference. As seen, the estimated C_n^2 values in the lower observation level (right-hand side plots of NAVSLaM

and CSAT3B) are distributed slightly to higher C_n^2 values than those in the lower observation levels (the other two left-hand side plots). The largest C_n^2 value at the upper observation level is about 10^{-12} $m^{-2/3}$, while at the lower observation level, the largest C_n^2 value exceeded 10^{-12} $m^{-2/3}$, and approaches 10^{-11} $m^{-2/3}$. Both NAVSLaM and CSAT3B indicated that the weakest turbulence (C_n^2) generally occurs when air temperature difference approaches zero. Under these conditions, NAVSLaM estimated the weakest C_n^2 value to be less than 10^{-19} $m^{-2/3}$, at the upper observation level. At the same time, CSAT3B shows a C_n^2 value not less than 10^{-16} $m^{-2/3}$. Again, in the case of the sonics, this is near their noise floors.

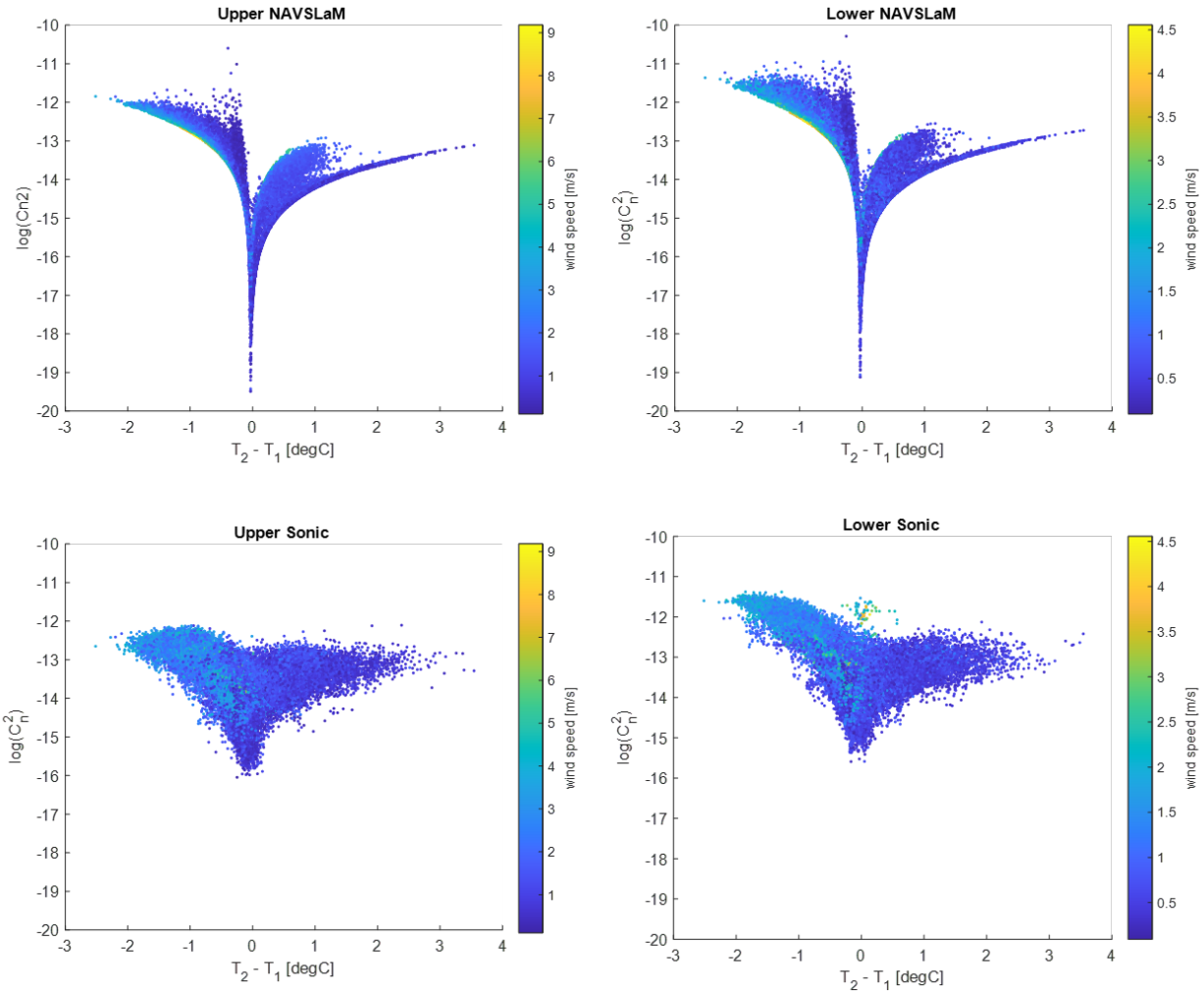


Figure 13. Plot of temperature differences versus log (C_n^2) predicted by NAVSLaM and measured by CSAT3B at two different heights.

To better understand how well NAVSLAM C_n^2 estimates agree with the CSAT3B measurements, we displayed the differences in the logarithm of these values against the mean temperature difference in Figure 14. The top plot correlates to the values of the upper sensors and the bottom to the lower sensors. A solid red line in both plots provide us an indicator when the NAVSLaM estimates perfectly agree with the CSAT3B measurements. In the region of the negative mean temperature differences, NAVSLaM estimates have considerably good agreement with the CSAT3B measurements, but it tends to slightly overestimate the C_n^2 values, as the temperature differences negatively increase as seen in the upper plot. This figure again proves that NAVSLaM underestimates the value of C_n^2 relative to the sonic observations when the temperature difference is close to zero and when the differences are positive. Furthermore, the agreement improves as the temperature differences positively increase.

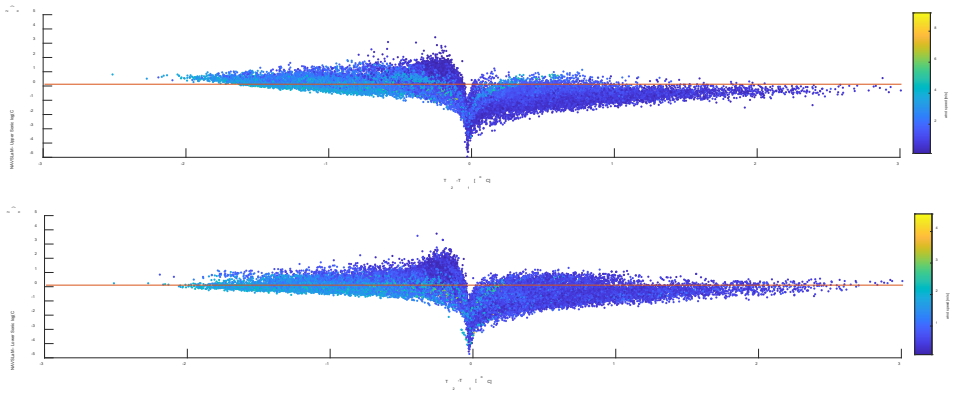


Figure 14. Differences between log (C_n^2) values estimated by NAVSLaM and corresponding values measured by CSAT3B sonic anemometers at two level heights.

V. CONCLUSION

This chapter is reproduced from M. Tamus, "Comparison of Optical Turbulence Prediction Models," M.S. Thesis, Naval Postgraduate School, Monterey, CA, USA, 2022 [1].

This study explored whether machine learning methods can predict optical turbulence using simple atmospheric parameters measured by more robust equipment. We conducted an experiment to measure point turbulence at two levels as well as to collect meteorological data such as wind speed, sonic temperature, air temperature, soil temperature, barometric pressure, and solar flux for three months at the Naval Research Laboratory Coastal Environmental Observation Station (NRL-CEOBS), which is located on the edge of Monterey Bay Beach near Watsonville. Using these simple atmospheric measurements and point turbulence measurement data, we developed an ML model to predict optical turbulence. From the results, we concluded that the ML regression analysis using a bagged ensemble of regression trees seems feasible for predicting turbulence at different heights. This model shows reasonable agreement with the observed turbulence values, regardless of the air temperature difference or other parameters that influence the amount of turbulence.

The same data set was used to predict optical turbulence using the two-level NAVSLaM model. The results show decent agreement with the observed turbulence values, particularly in unstable conditions, when the temperature at the lower level is greater than at the upper level. In stable conditions, when the temperature at the upper level is greater, NAVSLaM tended to slightly underestimate the turbulence values. In neutral conditions, when the temperatures of the two levels were approximately equal, NAVSLaM often dramatically underestimated the turbulence values; this is a known issue with MOST theory [10]. Some of the differences between the NAVSLaM predictions and the observed values of turbulence could be due to inhomogeneity caused by flow distortion.

Overall, the ML model appeared to work better than NAVSLaM for predicting the optical turbulence values that we observed. However, NAVSLaM is a more general model that should work reasonably well in a large variety of environments, whereas our ML

model is a product of the particular location where it was developed, and most likely would not work as well at a different location.

Although the results of this study show that machine learning can predict optical turbulence reasonably well, as a reminder, we evaluated the trained machine learning model with a set of test data collected in the same timeframe as the data used as training data. This research was supposed to collect data for a longer time; however, we could only process the data collected for about two months due to technical issues with the equipment. So, this trained ML model might be specific not only to the location but also to the period of the observations, which is from October to November.

There are still many opportunities to develop more general machine-learning models that apply to broader coastal environments. Some future studies include training a machine learning model with meteorological data collected throughout the year in the same place and testing the trained model with new meteorological data collected at different locations that have similar environmental characteristics. Besides that, training machine learning models with temperature and wind speed gradients or data derived from other meteorological parameters can be another consideration for future research.

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

[1] J. D. Ellis, *Directed-Energy Weapons: Promise and Prospects*. Washington, D.C, USA: Center for a New American Security, 2015. Accessed: May 20, 2022. [Online]. Available: https://www.cnas.org/publications/reports/directed-energy-weapons-promise-and-prospects

[2] D. H. Titterton, "Laser Beam Propagation," in *Military Laser Technology and Systems*, Norwood, MA, USA: Artech House, 2015, pp. 165–195.

[3] R. Mahon, C. I. Moore, M. S. Ferraro, W. S. Rabinovich, and P. A. Frederickson, "Comparison of maritime measurements of Cn2 with NAVSLaM model predictions," *Applied optics. Optical technology and biomedical optics*, vol. 59, no. 33, pp. 10599–10612, 2020, doi: 10.1364/AO.405185.

[4] P. A. Frederickson, S. Hammel, and D. Tsintikidis, "Measurements and modeling of optical turbulence in a maritime environment," *Proc. SPIE 6303*, vol. Atmospheric Optical Modeling, Measurement, and Simulation II, 630307, Sep. 2006, Accessed: May 12, 2022. [Online]. Available: https://calhoun.nps.edu/handle/10945/41322

[5] A. Sklavounos, "Measurements of Optical Turbulence and Analysis Using Machine Learning," M.S. Thesis, Naval Postgraduate School, Monterey, CA, USA, 2021.

[6] G. P. Perram, S. J. Cusumano, R. L. Hengehold, and S. T. Fiorino, "The Atmosphere," in *Introduction to Laser Weapon Systems*, 1st ed., J. S. Accetta, Ed. Albuquerque, NM, USA: Directed Energy Professional Society, 2010, pp. 203–232.

[7] R. L. Beland, "Propagation through Atmospheric Optical Turbulence," in *The Infrared and electro-optical systems handbook*, vol. Atmospheric Propagation of Radiation, 2 vols., J. S. Accetta, D. L. Shumaker, and F. G. Smith, Eds. Ann Arbor, MI, USA: Infrared Information Analysis Center, 1993, pp. 159–229.

[8] L. C. Andrews and R. L. Phillips, "Optical Turbulence in the Atmosphere," in *Laser Beam Propagation through Random Media*, Bellingham, WA, USA: SPIE, 2005, pp. 58–80. [Online]. Available: https://doi-org.libproxy.nps.edu/10.1117/3.626196.ch3

[9] C. Brennen, "An Internet Book on Fluid Dynamics," [Online]. Available: http://brennen.caltech.edu/fluidbook/Fluidbook.htm (accessed Oct. 10, 2023).

[10] A. S. Monin and A. M. Obukhov, "Basic laws of turbulent mixing in the surface layer of the atmosphere," *Tr. Akad. Nauk. SSSR Geophiz. Inst.*, vol. 24 (151), pp. 163-187., 1954.

[11] S. Barnett, J. Blau, P. Frederickson, and K. Cohn, "Measurements and Modeling of Optical Turbulence in the Coastal Environment," *Applied Sciences*, vol. 12, no. 10, pp. 4892–4903, 2022, doi: 10.3390/app12104892.

[12] IBM, "What is Machine Learning?," Jul. 06, 2022. https://www.ibm.com/cloud/learn/machine-learning (accessed Nov. 16, 2022).

[13] MATLAB, "Regression - MATLAB & Simulink." https://www.mathworks.com/help/stats/regression-and-anova.html?s_tid=CRUX_lftnav (accessed Oct. 27, 2022).

[14] G. Ciaburro, *MATLAB for Machine Learning: Functions, Algorithms, and Use Cases*. Birmingham, England: Packt, 2017. [Online]. Available:

https://ebookcentral.proquest.com/lib/ebook-nps/detail.action?pq-origsite=primo&docID=4987771
[15] S. Kumar, “Improving the Performance of Machine Learning Model using Bagging,” *Medium*, Jul. 04, 2020. https://towardsdatascience.com/improving-the-performance-of-machine-learning-model-using-bagging-534cf4a076a7 (accessed Nov. 14, 2022).
[16] M. Tamus, "Comparison of Optical Turbulence Prediction Models," M.S. Thesis, Naval Postgraduate School, Monterey, CA, USA, 2022

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California

3. Research Sponsored Programs Office, Code 41
   Naval Postgraduate School
   Monterey, CA 93943

4. Sarwat Chappell
   Office of Naval Research
   Arlington, VA 22203