To be published in Applied Optics:

Title: Machine-learning informed macro-meteorological models for the near-maritime environment

Authors: Christopher Jellen,Miles Oakley,Charles Nelson,John Burkhardt,Cody Brownellni

Accepted: 24 February 21

Posted 25 February 21

DOI: https://doi.org/10.1364/AO.416680

Machine-learning informed macro-meteorological models for the near-maritime environment

CHRISTOPHER JELLEN[1], MILES OAKLEY[2], CHARLES NELSON[2], JOHN BURKHARDT[1], CODY BROWNELL[1]

[1]*Mechanical Engineering Department, United States Naval Academy, 1 Wilson Rd., Annapolis, MD, 21402*
[2]*Electrical Engineering Department, United States Naval Academy, 1 Wilson Rd., Annapolis, MD, 21402*

Received XX Month XXXX; revised XX Month, XXXX; accepted XX Month XXXX; posted XX Month XXXX (Doc. ID XXXXX); published XX Month XXXX

Macro-meteorological models predict optical turbulence as a function of weather data. Existing models often struggle to accurately predict the rapid fluctuations in C_n^2 in near-maritime environments. Seven months of C_n^2 field measurements were collected along an 890 m scintillometer link over the Severn River in Annapolis, Maryland. This time series was augmented with local meteorological measurements to capture bulk-atmospheric weather measurements. The prediction accuracy of existing macro-meteorological models was analyzed in a range of conditions. Next, machine-learning techniques were applied to train new macro-meteorological models using the measured C_n^2 and measured environmental parameters. Finally, the C_n^2 predictions generated by the existing macro-meteorological models and new machine-learning informed models were compared for four representative days from the data set. These new models, under most conditions, demonstrated a higher overall C_n^2 prediction accuracy, and were better able to track optical turbulence. Further tuning and machine learning architectural changes could further improve model performance.

http://dx.doi.org/10.1364/AO.99.099999

1. INTRODUCTION

Turbulent mixing within the atmosphere causes rapid fluctuations in the local index of refraction resulting in optical turbulence. The intensity of this phenomenon is quantified by the index of refraction structure parameter, C_n^2. The magnitude of C_n^2 is directly related to the turbulent structure of the atmosphere which depends on both large-scale atmospheric forcing and on local environmental conditions. When a laser beam propagates along a path, the fluctuations in the local refractive index results in beam perturbations. These atmospheric effects can lower the energy received at a target or cause information loss in laser communication [1-3]. The degradation in laser beam quality and the resulting degradation in system performance can be estimated directly from predicted C_n^2 values.

At low altitudes, optical turbulence is caused predominantly by temperature gradients within the air [1-3]. Existing macro-meteorological models leverage local measurements of weather parameters to generate a C_n^2 prediction [4-7]. Air temperature, pressure, and humidity are of demonstrated importance in predicting optical turbulence [1-3,7]. Additionally, the wind speed and temporal hour weight are often included in explicit macro-meteorological models such as the Sadot macro-meteorological (Model 1), the Offshore-updated model in Wang (Model 2), and the ANOVA macro-meteorological model in Raj et. al (Model 3) [8-10]. These models have been applied in our near-maritime environment with mixed success [8-11].

In order to better understand the accuracy of these existing models, a new data set containing measurements of optical turbulence and the corresponding environmental parameters was collected over a near-maritime propagation path over the Severn River, Annapolis, MD. Prior investigations into using environmental parameters to predict the extent of optical turbulence suggests that machine learning and polynomial regression could be used to improve prediction accuracy [7,12]. To this end, the macro-meteorological Models 1 – 3 were evaluated on near-maritime data collected between January 1st 2020 and July 17th 2020. Additionally, this paper presents three new models developed on this data set and compares them to these existing models. Polynomial regression was applied, in addition to two machine-learning approaches which were suggested to be appropriate in [7]. The two machine learning models were the random forest model, and boosted regression tree model, whose structure is described further in [7,13-15]. The newly-trained and existing models were compared with measured C_n^2 using seven months of data gathered in the near-maritime environment of the Severn River, with four full days between June 16th 2020 and July 17th 2020 explored in detail.

2. THEORY

Random fluctuations in the atmosphere's index of refraction distort the wavefront of beams propagating through the medium. Temperature differences between pockets in the atmosphere called eddies contribute to these random fluctuations at optical frequencies. Pressure and humidity have also been shown to be important factors affecting optical turbulence [1, 2, 4, 7,12]. At low altitudes, especially in near-maritime environments, the refractive index of air can change rapidly in space and time, resulting in greater turbulent effects in the local atmosphere.

In order to predict the effect of atmospheric turbulence on beam propagation, the refractive index of the atmosphere can be expressed using Eq. (1) [1, 2]:

$$n(\hat{r}, t) = n_0 + n_1(\hat{r}, t) \quad (1)$$

where the local refractive index of the atmosphere n is defined to be the mean index of refraction for the non-turbulent atmosphere n_0 with the added turbulent fluctuations given by n_1, which is a function of both the position along the propagation path $\hat{r}$ and time t. By applying the assumptions of temporal homogeneity and normalizing the mean index of refraction n_0 to 1, the relationship in Eq. (1) no longer has time dependance. The resulting Eq. (2) is an approximation of the local refractive index of the atmosphere [1,2,4]:

$$n(\hat{r}) = 1 + n_1(\hat{r}) \quad (2)$$

The behavior of the refractive index is tied to the physical properties and interactions of the local atmosphere. Literature suggests that Eq. (2) can be modified for optical wavelengths; this relationship is expressed in Eq. (3) [2] as,

$$n(\hat{r}) \cong 1 + 79 \times 10^{-6} \frac{P(\hat{r})}{T(\hat{r})} \quad (3)$$

where the structure of the local index of refraction n is predicted from the pressure P in $[\mathrm{mBar}]$ and the temperature T in $[\mathrm{K}]$ along the path $\hat{r}$. Prior studies suggest that local fluctuations in pressure are generally negligible [2]. The random fluctuations in the index of refraction occur primarily due to temperature effects [2]. One way to describe the fine-scale characteristics of the temperature distribution is through the structure function for temperature. The outer eddy scale L_0 and an inner scale of l_0 define the size of large-scale and small-scale eddies in the propagation environment. The temperatures of neighboring points are correlated. The structure function calculates a difference in temperature at two points as a function of the proximity between the points. The flow is assumed to be locally homogeneous and isotropic, with a large separation of scales. The structure function for temperature is further assumed to depend only on the distance between points R, and on a constant, C_T^2, that characterizes the magnitude of the temperature differences in Eq. (4):

$$D_T(R) = \langle (T_1 - T_2)^2 \rangle = \begin{cases} C_T^2 l_0^{-\frac{4}{3}} R^2 & 0 \le R \ll l_0 \\ C_T^2 R^{2/3} & l_0 \le R \ll L_0 \end{cases} . \quad (4)$$

The concept of the structure function can be extended to describe local refractive index fluctuations, also with outer scale L_0 and inner scale l_0 in Eq. (5) [1,2]. The structure function describing the refractive index within statistically homogeneous isotropic turbulence is given by Eq. (5):

$$D_n(R) = \langle (n_1 - n_2)^2 \rangle = \begin{cases} C_n^2 l_0^{-\frac{4}{3}} R^2 & 0 \le R \ll l_0 \\ C_n^2 R^{2/3} & l_0 \le R \ll L_0 \end{cases} \quad (5)$$

where C_n^2 is the index of refraction structure constant, with units of $[m^{-\frac{2}{3}}]$. Using the relationship in Eq. (3), and the assumptions outlined in the formulation of Eq. (4) and Eq. (5), the constant C_n^2 can be calculated as in Eq. (6) [1, 2]:

$$C_n^2 = \left(79 \times 10^{-6} [\frac{K}{mbar}] \frac{P}{T^2}\right)^2 C_T^2 \quad (6)$$

where temperature T is measured in $[\mathrm{K}]$, pressure P is in $[\mathrm{mBar}]$, and C_T^2 has units of $[\mathrm{K}^2 \mathrm{m}^{-\frac{2}{3}}]$. The index of refraction structure constant C_n^2 is a measure of the intensity of the fluctuations in the refractive index of the atmosphere. The temperature structure parameter is estimated using point measurements of environmental parameters [7, 8]. In near-maritime environments, a wide range of environmental parameters may be required to accurately predict the intensity of optical turbulence.

In addition to point-wise measurements of the medium, the path-averaged refractive index structure parameter is determined from optical measurements at a target. The variance in beam irradiance are measured and normalized to generate a scintillation index σ_I^2, defined in Eq. (7) [1, 2, 4, 5, 7, 12]:

$$\sigma_I^2 = \frac{\langle I^2 \rangle - \langle I \rangle^2}{\langle I \rangle^2} \quad (7)$$

where I is the irradiance of the beam at the target at a given time under the assumption of time-averaged ergodicity [1,2]. The scintillation index is used to calculate C_n^2 for a plane-wave using the relationship in Eq. (8) [1, 2]:

$$\sigma_I^2 = 1.23 C_n^2 k^{\frac{7}{6}} L^{\frac{11}{6}} \quad (8)$$

where the path length L and wave number k of a beam are related to the observed scintillation index through the structure constant C_n^2. Experimentally, this measurement is taken using a scintillometer, a device which records variance in irradiance of a known beam on its receiver positioned a distance L from its transmitter.

Estimating C_n^2 directly can be achieved through use of Eq. (6). This process relies on knowledge of local temperature within each of the eddies along the path. Temperature measurement of large and small scale eddies along the propagation path allow for approximations for the temperature structure function [1,2]. Rather than measuring the absolute temperature difference between different eddies directly, bulk atmospheric similarity theories employ the use of mean flow characteristics along the propagation path. Monin–Obukhov similarity (MOS) theory is used to describe flow behavior for use in C_n^2 prediction [1-3,8]. MOS theory calculates average properties along a path by assuming parameters such as temperature, humidity, and wind speed are path-invariant. Using MOS theory, C_n^2 is predicted using a combination of temperature and humidity effects, where temperature and humidity are modeled with structure parameters from bulk weather measurements [1-3].

The assumption of path invariance allows for approximation of the temperature and humidity structure parameters using point measurements of environmental parameters. And while these assumptions simplify the process of predicting C_n^2, they could contribute to the overestimation of the intensity of optical turbulence in low-altitude maritime environments [5-8]. In these environments, a wider range of environmental parameters than those used in MOS theory may be required to accurately predict the intensity of optical turbulence.

3. DATASET

A. C_n^2 and Environmental Parameter Measurement

The core of this study is an expansive data set consisting of time-series C_n^2 observations and measurements of local meteorological parameters. These measurements were obtained over an 890 $[\mathrm{m}]$ propagation path established over the Severn River at the United States Naval Academy (USNA) in Annapolis, Maryland. The propagation environment is bounded behind each side of the scintillometer link by land. The Severn River is a tidal estuary, which opens into the Chesapeake Bay. Optical turbulence was measured as C_n^2, captured with a Scintec BLS450 scintillometer each minute. These readings were used to train new models as well as to compare model predictions. The beam propagation path is over the river, 3.1 $[\mathrm{m}]$ above mean low tide. The mean diurnal range at this location is 0.4 $[\mathrm{m}]$, indicating that the scintillometer's height above the water line changed slowly and had a limited range [16]. While the propagation path was entirely over water, large land masses behind the transmitter and receiver impacted the propagation environment. As the local atmosphere is affected by the land masses on either side of the path, the link is considered to be in a near-maritime environment [6,7,9,12]. Readings captured by the BLS450 scintillometer are accompanied by error codes where applicable. Measured scintillation in the low $10^{-16}[\mathrm{m}^{-\frac{2}{3}}]$ could possibly be in the noise based on analysis from the Scintec planning tool. During periods of high beam scintillation above mid 10^{-12} $[\mathrm{m}^{-\frac{2}{3}}]$ or link interruptions from boat traffic, the scintillometer records errors [17]. All measurements with applicable error codes were removed from the dataset, but represented less than 2% of observations between January 1st 2020 and July 17th 2020. After removing scintillometer observations with measurement errors, there were no missing values in the C_n^2 data set.

A diagram of the scintillometer link location, along with the receiver and transmitter, is given in Fig. (1).

Fig. 1. Beam propagation path with receiver (left) and transmitter (right) for BLS 450 scintillometer across the Severn River in Annapolis, Maryland [18].

Optical turbulence measurements collected using the scintillometer link in Fig. (1) was augmented by local meteorological parameters collected using a weather station near the scintillometer link. The meteorological measurements were averaged over 10-minute intervals. Two meteorological data sources, a Davis VANTAGE PRO2

weather station and a data buoy maintained by the National Data Buoy Center (NDBC), provided weather measurements for the data set [19,20].

The Davis weather station is situated next to the receiver of the scintillometer link, no more than $5\ [m]$ away, at roughly the same elevation over the water line. This data source recorded measurements of air temperature, dew point temperature, pressure, humidity, wind speed, wind direction, gust wind speed, gust wind direction, solar radiation, UV index, and air density from January 1st 2020 until July 17th 2020.

There were few missing values in the set of meteorological measurements from the local weather station. While the study initially attempted to record rainfall, the weather station failed to accurately measure the amount of rain over each 10-minute time interval. In late February, the sensor used to measure the UV index was re-centered to ensure accurate measurement. A diagram of the USNA weather station configuration, with labeled sensors, is provided in Fig. (2).

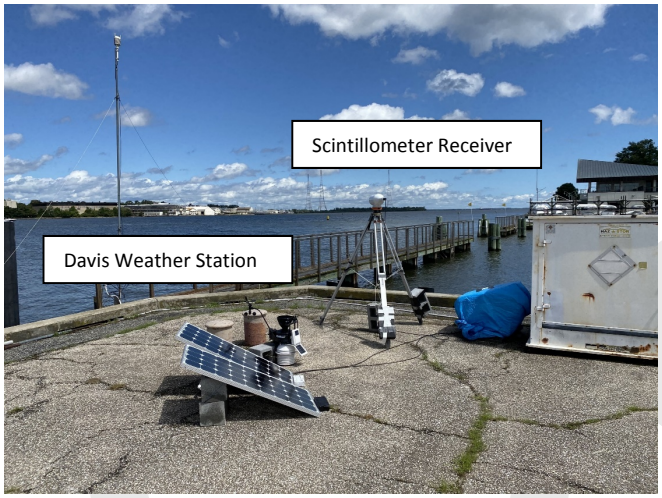


Fig. 2. USNA weather station and scintillometer receiver configuration.

The weather station in Fig. (2) and the scintillometer link in Fig. (1) provided the core of the data set. The atmospheric parameters were collected approximately $1\ [\mathrm{m}]$ below the scintillometer measurements. In addition to the meteorological parameters collected locally, measurements taken from a National Data Buoy Center (NDBC) station were used to measure water temperature. The NDBC station at Thomas Point Light, in the vicinity of the scintillometer link, was used to collect 10-minute measurements of water temperature between January 1st 2020 and July 17th 2020 [20].

Two additional temporal parameters were added to the data set before final standardization and analysis. In an attempt to capture the potential relationship between optical turbulence and sunlight, the temporal hour was calculated for each observation of C_n^2 [7, 12, 21]. Temporal hour weight is obtained by dividing the difference between the time of each observation and local sunrise time by 1/12 the time between local sunrise and sunset, as in Eq. (9) [8]:

$$th_W = \frac{12(t_{observation} - t_{sunrise})}{t_{sunset} - t_{sunrise}}\ . \qquad (9)$$

The recorded time of each observation was determined using the scintillometer internal clock, while local sunrise and sunset were recorded by the USNA Oceanography Department sunrise data set [7, 12, 21]. Additionally, the daylight duration was measured as the time between sunrise and sunset for each day using Eq. (10):

$$S = t_{sunset} - t_{sunrise} \qquad (10)$$

where *S* represents the time duration between sunrise and sunset for each day. The parameter *S* was included in an attempt to capture seasonal dependence within the dataset.

B. Dataset Preparation

The dataset served both to evaluate the efficacy of existing macro-meteorological models for predicting optical turbulence in the near-maritime environment as well as to train new models. The first step in synthesizing a complete data set containing C_n^2, meteorological, and oceanographic measurements was time-series standardization. The data set contained $23{,}409$ observations between January 1st 2020 and June 15th 2020, and 4,590 observations between June 15th 2020 and July 17th 2020. The data from June 15th 2020 to July 17th 2020 was used as a validation set to compare the relative prediction accuracy of existing macro-meteorological models and the new models trained specifically for the near-maritime environment. The data from January 1st 2020 to June 15th 2020 was divided into a training set of 80% of observations, and a test set of 20% of observations using random selection. The training and test sets were used primarily to develop machine-learning informed macro-meteorological models. They were also used to evaluate current explicit macro-meteorological models. More information in the data set preparation considerations is available in [12]. Summary statistics for 7 of the environmental parameters collected in the data set are given in Table 1.

Table 1. Data Set Summary

	Air Temperature [°C]	Relative Humidity [%]	Wind Speed $\left[\frac{m}{s}\right]$	Pressure [mBar]	Solar Radiation $\left[\frac{W}{m^2}\right]$	Air-Water Temperature Difference [°C]	Solar Hour
Mean	14.3	69.9	1.4	1018	176.5	0.0	5.6
Std.	9.1	18.9	1.0	8.1	265.2	3.3	7.0
Minimum	-3.4	14.0	0.0	986.1	0.0	-10.6	-7.0
25%	7.3	55.0	0.4	1012.7	0.0	-2.0	0.0
50%	12.4	71.0	1.3	1017.5	11.0	-0.1	6.0
75%	21.7	86.0	1.8	1023.3	281.0	1.7	12.0
Maximum	38.1	98.0	7.6	1044.0	1214.0	14.7	19.0

In Table 1, the mean measurement, standard deviation of measurements, minimum, maximum, and interquartile ranges for observations are reported for each environmental parameter column. These summary statistics were used to validate the measurements collected for the dataset. Notably, the air-water temperature difference was typically between -2 [°C] and 2 [°C], but had a large standard deviation relative to the range of the middle 50% of observations. The maximum and minimum values are also notable, as they corresponded to rapid changes in air temperature. The full set of raw measurements collected from the scintillometer and local weather station are publicly available at the repository in ref. [22]

C. Measurement Distribution Analysis

Following dataset preparation, cleaning, and time standardization, the measurements of each parameter of interest were analyzed and compared with existing literature. The first parameter of interest was optical turbulence as measured by C_n^2. The distribution and central tendency of C_n^2 measurements from each month between January 2020 and July 17th 2020 are reported as histograms in Fig. (3).

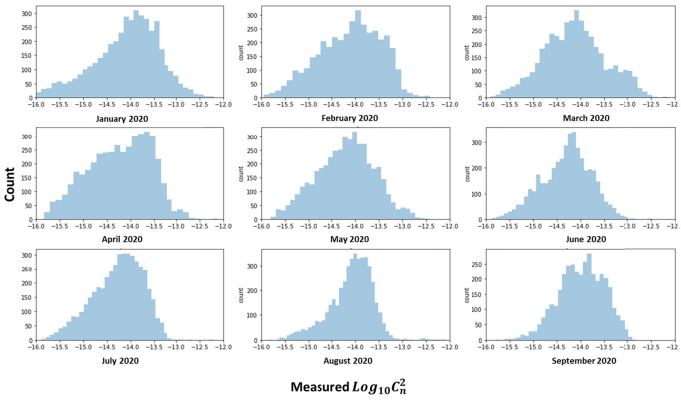


Fig. 3. Monthly histograms of counts of observed C_n^2 from January 2020 to July 17th 2020.

The monthly distributions of C_n^2 measurements were in-line with existing literature for the near-maritime environment [4-7]. The extent of optical turbulence in the low-altitude near-maritime environment is generally greater than for open-ocean propagation. The histograms in Fig. (3) indicate that optical turbulence may generally be higher during periods of greater temperature change, such as early spring [4-7].

After exploratory data analysis and validation of scintillometer readings, the data set was used to analyze three common models from literature, as well as train new polynomial regression and machine-learning based models for C_n^2 prediction. The models of interest from the literature were Model 1 (Sadot Macro-meteorological Model in ref. [8]), Model 2 (Wang's Offshore-updated Macro-meteorological Model in ref. [9]), and Model 3 (Raj's Analysis of Variance, or ANOVA Model in ref. [10]). These models were selected because they did not include height or propagation surface as parameters. Measurements were used to predict both daytime and nighttime measures of C_n^2 [8-11]. Based on prior work in [7,12], a polynomial regression model was trained to fit the measurements in the data set between January 1st 2020 and June 15th 2020. Finally, the random forest model and boosted regression tree models were selected as appropriate machine-learning based models for predicting C_n^2 from meteorological parameters based on the work in ref. [7] and ref. [12].

4. MACRO-METEOROLOGICAL MODEL EVALUATION

The macro-meteorological models evaluated in this study make use of measured meteorological parameters to generate predicted values of C_n^2. Prior literature has investigated the relationship between meteorological parameters such as the air temperature, relative humidity, wind speed, and temporal hour weight on optical turbulence. Due to variation between local propagation environments, these models typically assume some explicit structural relationship between meteorological parameters and the extent of optical turbulence, and then use local measurements to fit a regression model using the assumed structure. Polynomial regression combined with in situ measurements of optical turbulence and meteorological parameters was used to develop Model 1 [8]. Using a new data set of measurements for a coastal propagation environment, this model was updated in Model 2 [9]. Finally, a more general model which attempted to capture seasonal variations over a range of locations was trained as Model 3 [10]. The prediction accuracy of each of these three models for data collected over the Severn River offers insight into the wider applicability of macro-meteorological models in near-maritime environments.

A. Sadot Macro-meteorological Model (Model 1)

Model 1 uses macro scale measurements of air temperature, relative humidity, wind speed, and temporal hour weight to generate a prediction for C_n^2. This model has been widely applied in [8-11]. In the near-maritime environment investigated in [9], the model was found to generally over-estimate the extent of optical turbulence. The explicit form of Model 1 is given in Eq. (11):

$$C_n^2 = (3.8 \times 10^{-14})W + f(T) + f(U) + f(RH) - (5.3 \times 10^{-13}),$$

where

$$f(T) = (2.0 \times 10^{-15})T$$
$$f(U) = (-2.5 \times 10^{-15})U + (1.2 \times 10^{-15})U^2 - (8.5 \times 10^{-15})U^3$$
$$f(RH) = (-2.8 \times 10^{-15})RH + (2.9 \times 10^{-17})RH^2 - (1.1 \times 10^{-19})RH^3. \quad (11)$$

In Eq. (11), W denotes the temporal hour weight, as in [8], T denotes the temperature in $[K]$, RH denotes the relative humidity in $[\%]$, and U denotes the wind speed in $[\frac{\mathrm{m}}{\mathrm{s}}]$. Comparing the predicted C_n^2 with the observed C_n^2 gives insight into the applicability of this model in the near-maritime environment. These predictions were made on atmopsheric data collected at an elevation of about 2 [m], and plotted against C_n^2 measured at an elevation of 3.1 [m] over the mean low tide.

Using data gathered between January 1st 2020 and July 17th 2020, Model 1 was used to predict C_n^2 and compared the value of C_n^2 measured by the scintillometer for each 10-minute interval. The accuracy of the model in the near-maritime environment of the Severn River is visualized in Fig. (4).

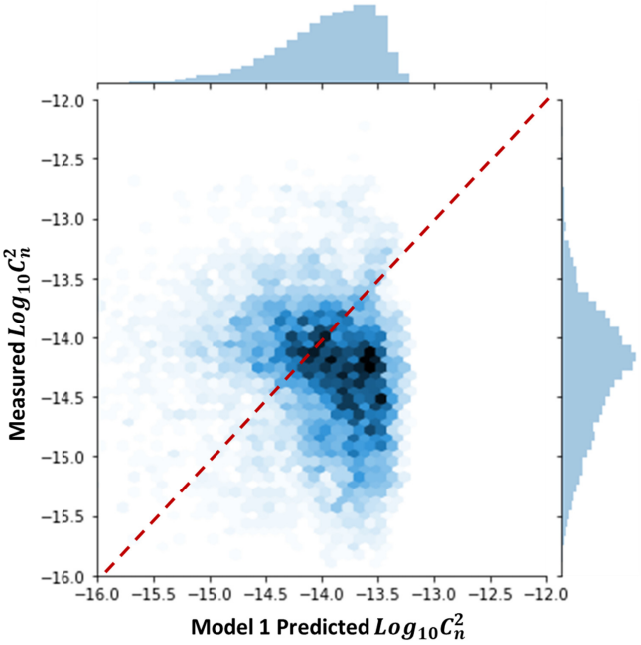


Fig. 4. Joint distribution predicted and measured $Log_{10}C_n^2$ for Model 1 in ref. [8] between January 2020 and July 2020 where the red line indicates perfect model performance.

In Fig. (4), Model 1 tends to over-predict the extent of optical turbulence. The mean average error for the Macro-meterological model on the Severn River data set was 1.67×10^{-14} $[\mathrm{m}^{-\frac{2}{3}}]$. The distribution of the predicted $Log_{10}C_n^2$ for the Macro-meterological model has a left skew, and most predictions fall between 1×10^{-15} and 5×10^{-14} $[\mathrm{m}^{-\frac{2}{3}}]$. This trend in model prediction was also observed in the coastal environment around Yantai, China in [9]. Due to the general over-estimation of C_n^2 in that study, an updated Offshore Macro-meterologiucal model was developed by Wang.

B. Wang Offshore Macro-meteorological Model (Model 2)

Model 2 in [9] uses the same environmental parameters as the Model 1 in [8] to generate predictions of C_n^2, with different coefficients. The explicit form of the Macro-meteorological model is given in Eq. (12):

$$C_n^2 = (-1.58 \times 10^{-15})W + f(T) + f(U) + f(RH) - (7.44 \times 10^{-14}),$$

where

$$f(T) = (2.74 \times 10^{-16})T$$
$$f(U) = (3.37 \times 10^{-16})U + (1.92 \times 10^{-16})U^2 - (2.8 \times 10^{-17})U^3$$
$$f(RH) = (8.3 \times 10^{-17})RH - (2.22 \times 10^{-18})RH^2 + (1.42 \times 10^{-20})RH^3. \quad (12)$$

In Eq. (12), W denotes the temporal hour weight using [8], T denotes the temperature in [K], RH denotes the relative humidity in [%], and U denotes the

wind speed in $[\frac{\mathrm{m}}{\mathrm{s}}]$. Comparing the predicted C_n^2 with the observed C_n^2 gives insight into the applicability of this model in the near-maritime environment. These predictions were made on atmopsheric data collected at an elevation of about 2 $[\mathrm{m}]$, and plotted against C_n^2 measured at an elevevation of 3.1 $[\mathrm{m}]$ above mean low tide.

Predicted C_n^2 is compared to the values of C_n^2 measured by the scintillometer for each 10-minute interval in Fig. (5).

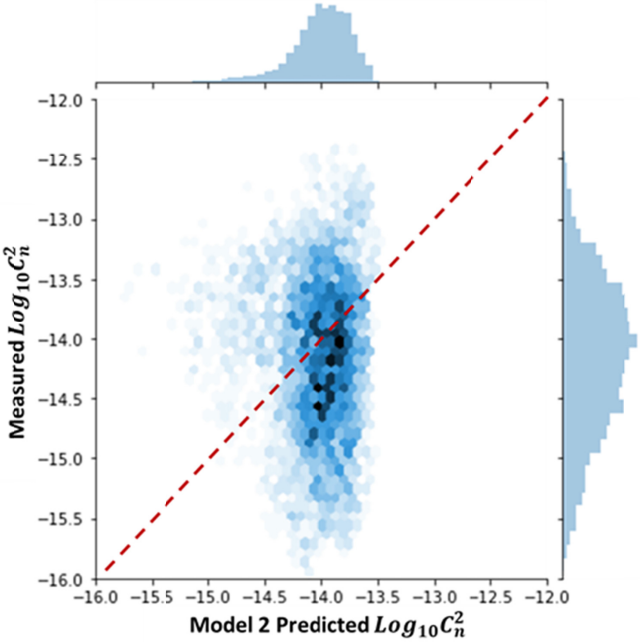


Fig. 5. Joint distribution predicted and measured $Log_{10}C_n^2$ for Model 2 in ref. [9] between January 2020 and July 2020 where the red line indicates perfect model performance.

In Fig. (5), the Offshore Macro-meterological model in [9] captures the mean observed value of C_n^2 more effectively than the model in [8] for the Severn River environment. The model's prediction distribution had narrow tails when compared to the distribution of C_n^2 field measurements. The mean average error for the Offshore macro-meterological model on the Severn River data set was 1.61×10^{-14} $[\mathrm{m}^{-\frac{2}{3}}]$. When using the weather parameter measurements in this data set, most predictions fall between 5×10^{-15} and 1×10^{-14} $[\mathrm{m}^{-\frac{2}{3}}]$. While the overall error of Model 2 is lower than that of Model 1, it fails to capture variation in observed optical turbulence effectively.

C. Raj ANOVA Model (Model 3)

Model 1 and Model 2 used measurements from one location under specific conditions to derive coefficients for environmental parameter regression. The team in Raj et. al. attempted to gather C_n^2 observations under a wider variety of atmospheric conditions by measuring optical turbulence in different locations over the course of one full year [10]. Data was collected at four locations simultaneously in an effort to develop a new model using the ANOVA approach, which analyzes differences in sample means and generates estimation models from the sample data. The first model from this effort minimizes the number and order of parameters needed to achieve a high R^2 value for the mixed location data set. The explicit form of Model 3 is given in Eq. (13):

$$\begin{aligned} C_n^2 = 1 \times 10^{-14}(399.774 &+ (4.88372)U - (14.7804)T - (5.58958)RH \\ &- (2.32469)U^2 + (0.203989)UT - (0.0616557)URH \\ &+ (0.153508)TRH + (0.170863)U^3 + (0.00276336)T^3 \\ &+ (0.000116384)TRH^2 \\ &+ (9.30618 \times 10^{-5})RH^3). \end{aligned} \quad (13)$$

In Eq. (13), U denotes the measured wind speed in $[\frac{\mathrm{m}}{\mathrm{s}}]$, T denotes the temperature in [°C], and RH denotes the relative humidity in [%]. Unlike Eq. (11) and (12), Model 3 does not make use of the temporal hour weight as a modeling parameter. The order of environmental parameters and their interaction terms is also notably higher than in previous models.

As with the prior two models, the predicted C_n^2 is compared against the measured C_n^2 captured bwteeen January 1st 2020 and July 17th 2020. The prediction accuracy is presented visually in Fig. (6).

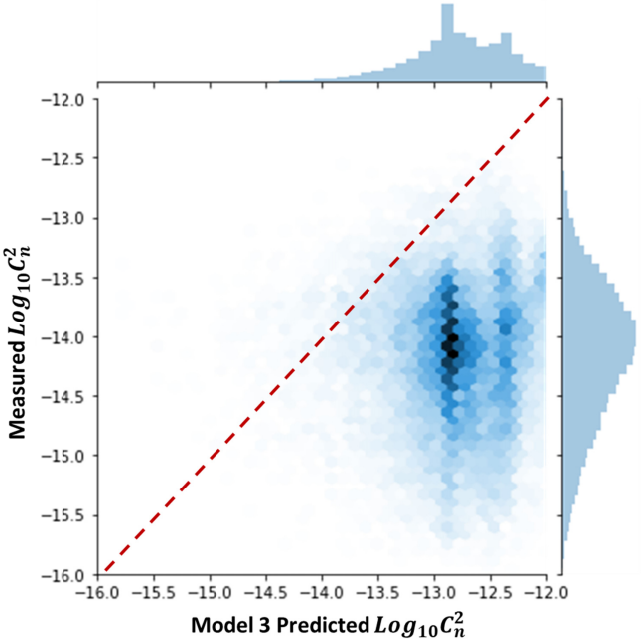


Fig. 6. Joint distribution predicted and measured $Log_{10}C_n^2$ for Model 3 in ref. [10] between January 2020 and July 2020 where the red line indicates perfect model performance.

Although Model 3 was developed to better capture general meteorological dependencies for predicting optical turbulence, the conditions under which the data was gathered differed materially from the near-maritime environment of the Severn River. The low end of the air temperature range for the Severn River data was much lower than that in Raj et. al., further, the relative humidity was consistently higher in the near-maritime environment than for some locations in Raj et. al. The prevalence of predicted C_n^2 above 5×10^{-13} $[\mathrm{m}^{-\frac{2}{3}}]$ could be attributed to the consistently high humidity between April 2020 and July 2020. Although the range of predicted C_n^2 was larger than for Model 1 or Model 2, the trend of over-prediction in the extent of optical turbulence was more pronounced. The mean average error for Model 3 on the Severn River data set was 2.68×10^{-13} $[\mathrm{m}^{-\frac{2}{3}}]$. While Model 1 and Model 2 were unimodal, Model 3 appeared to be bimodal. The predicted C_n^2 were typically either just above 1×10^{-13} $[\mathrm{m}^{-\frac{2}{3}}]$ or around 5×10^{-13} $[\mathrm{m}^{-\frac{2}{3}}]$.

The general trend to over-predict C_n^2, and failure to capture variance, for the near-maritime environment of the Severn River seen in Models 1, 2 and 3 motivated the development of new models to more accurately predict C_n^2. The first of these models used linear regression, with the environmental parameters included as predictors limited using a similar ANOVA approach to that applied in Raj et. al. [10]. Next, machine-learning informed regression tree based models were trained on the January to June section of the Severn River data set, and evaluated on the July 2020 data.

5. MACHINE-LEARNING MODELS

Machine-learning techniques were applied to generate new understanding of potential physical relationships between environmental parameters and optical turbulence. A random forest and boosted regression tree model were developed, in addition to a polynomial regression model. Observations from January 1st 2020 through June 15th 2020 were used to train and test models, with 20% of the observations in this range randomly selected for the test set. Next, observations from

June 16th 2020 until July 17ht 2020 were used as a validation set. This allowed some evaluation of the forecasting potential for these new models.

The data set used for training and testing contained observations of C_n^2 and the corresponding measurements of seven distinct environmental parameters. Prior literature has investigated a larger set of predictor variables [7]. Although greater numbers of predictor variables may improve short-term prediction accuracy, the capacity for model forecasting was limited [7]. The six environmental parameters selected to train the new models are given under Table 2.

Table 2. Environmental Parameters Used for Model Training

Parameter	Units	Data Source
Air Temperature	[°C]	Weather Station
Air-Water Temperature Difference	[°C]	Weather Station, Data Buoy
Relative Humidity	[%]	Weather Station
Wind Speed	$[\frac{\mathrm{m}}{\mathrm{s}}]$	Weather Station
Solar Radiation	$[\frac{\mathrm{W}}{\mathrm{m}^2}]$	Weather Station
Pressure	[mBar]	Weather Station

The six environmental parameters used were air temperature in [°C], relative humidity in [%], wind speed in $[\frac{\mathrm{m}}{\mathrm{s}}]$, pressure in [mBar], solar radiation in $[\frac{\mathrm{W}}{\mathrm{m}^2}]$, and air-water temperature difference in [°C]. The height above mean low tide is not included as a feature. Some features such as air temperature, pressure, and humidity have height dependence. Over the diurnal range of 0.4 [m], as discussed in Section 3 A the effect is assumed to be minimal. The set of measurements of these environmental parameters was not normalized before training. This may allow the models to better capture seasonal changes in the local environment, and aide in forecasting ability, but could potentially limit the range of propagation environments for which these models provide accurate predictions of C_n^2.

A. Linear Model with Polynomial Regression

To begin, first-order and second-order polynomial models were trained to determine the potential efficacy of general linear models in generating predictions for C_n^2. When training a polynomial model, both the training and the test set were combined and parameters were calculated to minimize the sum of square errors. The training scheme was similar to the ANOVA approach applied in [10]. Although the volume of data available for training was only about 25% of that in [10], the 9 months of 10-minute averaged data was more temporally expansive than some prior studies [6-8]. The explicit form of the polynomial regression model trained on the data between January 1st 2020 and June 15th 2020 is given in Eq. (14):

$$\log_{10} C_n^2 = -(0.102)T - (0.049)RH - (2.32469)\mathrm{U} - (0.80)P + (0.44)SR + (0.00487)AWT + (1.59)SH + (0.184)TRH. \quad (14)$$

In addition to the measurements of air temperature T in [°C], relative humidity RH in [%], and wind speed U in $[\frac{\mathrm{m}}{\mathrm{s}}]$ used by Models 1, 2, and 3, this model incorporated measurements of solar radiation SR in $[\frac{\mathrm{W}}{\mathrm{m}^2}]$ from [8], air-water temperature difference AWT in [°C], solar hour SH in [Hr], and atmospheric pressure P in [mBar]. Possibly because this model was developed on locally-collected environmental measurements, the prediction error was notably smaller than for Models 1, 2, and 3. In contrast to the existing macro-meterological models evaluated on the data set, the measured and predicted values of $Log_{10}C_n^2$ are compared only between June 16th 2020 and July 17th 2020 in Fig. (7). These measurements were not used to develop the model's coefficients.

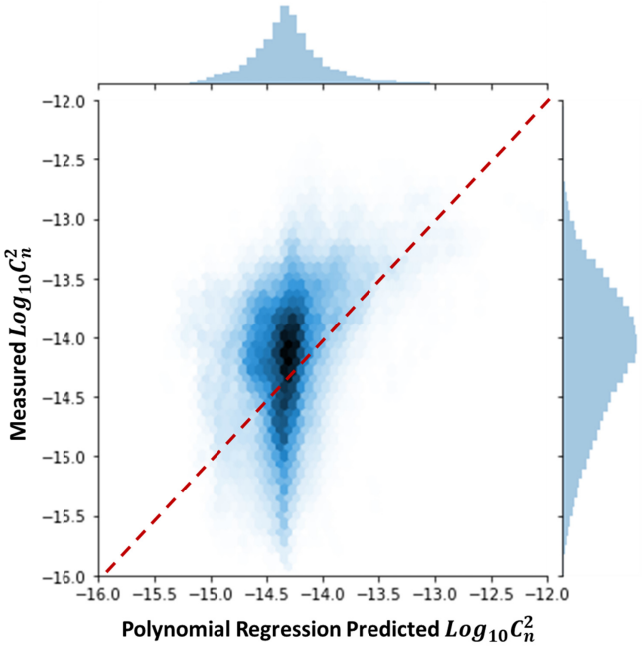


Fig. 7. Joint distribution predicted and measured $Log_{10}C_n^2$ for the Polynomial Regression model test data between January 1st 2020 and July 17th 2020.

The polynomial regression model trained on the Severn River data set better captured the shape of the observed $Log_{10}C_n^2$ distribution than the other explicit models, although it failed to capture the longer tails of the actual distribution. The model captured the mean level of optical turbulence fairly well, and compared favorably with other explicit models for predicting optical turbulence above $5 \times 10^{-13}[\mathrm{m}^{-\frac{2}{3}}]$. The model failed to accurately predict optical turbulence on the low-end, rarely predicting a value of C_n^2 below 5×10^{-14} $[\mathrm{m}^{-\frac{2}{3}}]$. The mean average error for the linear model trained with polynomial regression on the Severn River data set was 5.47×10^{-15} $[\mathrm{m}^{-\frac{2}{3}}]$, much lower than for any model previously investigated in this study.

Analysis of the prediction accuracy of the polynomial regression model informed the choice to train machine-learning based models based on regression trees. The regression tree approach has been applied to similar physical problems in [7,13-15], as well as to optical turbulence prediction in [7]. The linear model architecture which results from polynomial regression may be less appropriate than a non-linear model architecture in capturing the underlying physics which result in optical turbulent effects [7].

B. Random Forest Model

A regression tree is a nonparametric model which seeks to divide the data set into a number of distinct, non-overlapping regions in order to generate a prediction for a response variable given an input predictor observation [13, 14]. In this paper, regression trees will seek to partition $\mathbb{R}^7$, such that given some observation $\boldsymbol{x}^{[i]} = x_1^{[i]}, x_2^{[i]}, x_3^{[i]}, \ldots, x_7^{[i]}$ excluded from the training set, there exists a prediction $\hat{y}^{[i]}$ that depends only on the region in which $\boldsymbol{x}^{[i]}$ lies. More information on regression tree structure is available in [7,13,14].

As a model, each regression tree takes the form of a nonparametric model from $\mathbb{R}^7$ to $\mathbb{R}$. While a linear model on $\mathbb{R}^7$ is smooth, a regression tree surface can be jagged [15]. This may enable the regression tree model to better capture the physics associated with near-maritime optical turbulence. A simple regression tree example for predicting C_n^2 from measurements of environmental parameters is given in Fig. (8).

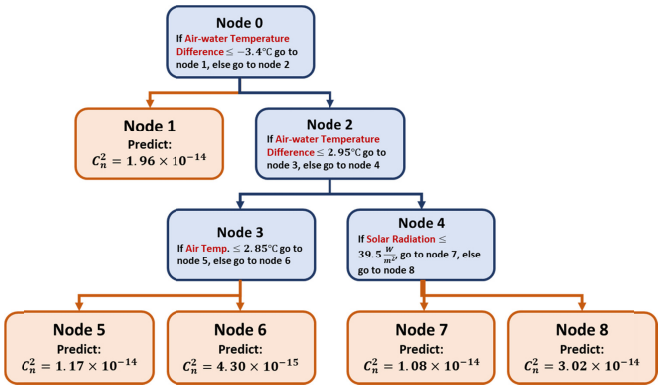


Fig. 8. Sample regression tree model for predicting C_n^2.

In Fig. (8), each rectangle is called a node, with the values for C_n^2 in each terminal node used as the prediction for all measurements which satisfy the conditions of its parent node. As an example, for any measurement with an air-water temperature difference below -3.4 [°C], the predicted C_n^2 for the regression tree in Fig. (8) is $1.96 \times 10^{-14} [\mathrm{m}^{-\frac{2}{3}}]$. Each of the decision boundaries in a parent node are determined by the training data used to develop the regression tree. The model in Fig. (8) is the 5-terminal-node regression tree with the lowest mean average error based on the training data.

In addition to explicit parameters, similar to coefficients in a linear model, machine-learning models include hyper-parameters. These hyper-parameters define the model architecture, and are not based on measured data. Examples include the depth of each regression tree, and the number of parameters considered. In a polynomial regression model, the highest order polynomial could be considered a hyper-parameter. In order to train the regression tree based models, a variety of potential values for hyper-parameters were evaluated. Those hyper-parameters which minimized prediction error on the test set were used as the final values for the machine-learning informed macro-meteorological models. Using a wide grid search, hyperparameters including the number of splits within each tree and maximum depth were investigated. The tuning process suggested that at least 10 observations were required for each regression tree split and the maximum depth should not exceed 16 nodes.

For the random forest architecture, the number of distinct regression trees used in the ensemble and number of predictors were most relevant in improving prediction accuracy. Tuning suggested that 500 distinct trees with a random subset of four of the seven available environmental parameters was appropriate to maximize test-set prediction accuracy. More information on regression-tree based model training is available in [7,13,14]. A visual representation of a random forest, made up of many distinct regression trees, is given in Fig. (9).

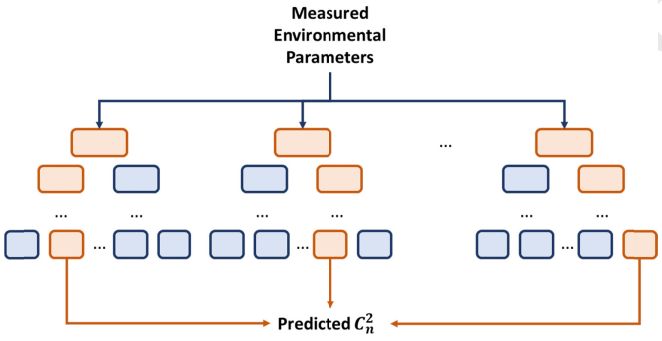


Fig. 9. Visual representation of a random forest model, where the individual environmental parameter measurements match the conditions of the orange nodes for each tree, and the average of all orange terminal nodes is used as the predicted C_n^2.

In Fig. (9), many individual regression trees were trained on a subset of the available environmental parameters, using the training data set. The simple average of predictions generated by these indiviudal trees is used as the model prediction. The decision boundary within each node is determined to miminize a given error metric on the available data. The same error metrics used to evaluate current macro-meterological models, as well as the linear model trained using polynomial regression, are applicable to the machine-learning informed models. As with the linear model, measured and predicted $Log_{10}C_n^2$ are compared between June 16th 2020 and July 17th 2020 in Fig. (10).

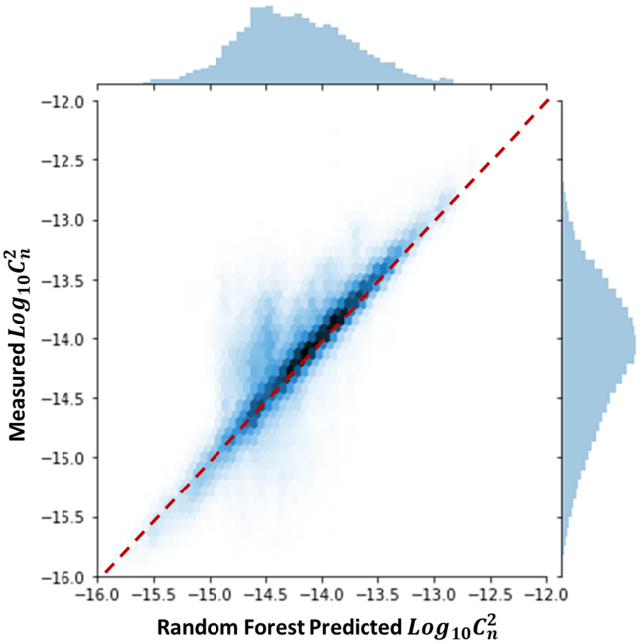


Fig. 10. Joint distribution predicted and measured $Log_{10}C_n^2$ for the random forest model test data between January 1st 2020 and July 17th 2020.

The distribution of predicted $Log_{10}C_n^2$ had both the long tails and uniform shape of the measured $Log_{10}C_n^2$ on the validation set. As with the polynomial regression model, observations from the validation set were never used in model training or tuning, to better capture the prediction accuracy and forecasting potential of the random forest model. The mean average error on this validation data set was $5.35 \times 10^{-15} \left[\mathrm{m}^{-\frac{2}{3}}\right]$. This performance indicated that machine-learning informed models could be successfully applied in the near-maritime environment. Notably, the random forest model captured the mean, high, and low observed optical turbulence well. There was some tendency to over-estimate the extent of optical turbulence below 5×10^{-15} $[\mathrm{m}^{-\frac{2}{3}}]$, but this was less extensive than for all prior models. As the random forest model was trained and tuned on data between January 1st 2020 and June 15th 2020, the high prediction accuracy on the data captured between June 16th 2020 and July 17th 2020 indicates remarkable forecasting accuracy.

While random forest architectures are widely applied in some physical domains, other competing non-parametric architectures could potentially be more appropriate. One such architecture is the boosted ensemble regression tree, which leverages the same non-parametric regression trees as the random forest model in a distinct way [7,13-15].

C. Boosted Regression Tree Model

As with the random forest model, the boosted regression tree model used a training set and test set to find the optimal internal weights for model parameters. The hyperparameters which defined the underlying regression tree architecture were similar to those in the random forest model. The tree depth, minimum number of observations for each split, and number of distinct regression trees were selected using a grid search over a wide range of possible values. The boosted model was

distinct from the random forest model in that its constituent regression trees were not equally weighted in the final model. Each subsequent regression tree was trained to minimize the negative gradient of the error function as in Eq. (15):

$$\overline{\boldsymbol{y}}_i = -\frac{\partial\left(\frac{1}{2}\left(\boldsymbol{y} - F_i(\boldsymbol{X})\right)^2\right)}{\partial F_i(\boldsymbol{X})}$$

where,

$$F_i(\boldsymbol{X}) = F_{i-1}(\boldsymbol{X}) + \lambda\left(T_i(\boldsymbol{X})\right). \quad (15)$$

In Eq. (15), $\overline{\boldsymbol{y}}_i$ denotes gradient in model error in the i^{th} step in the model given by $F_i(\boldsymbol{X})$, where i indicated the number of trees which have been incorporated into the model at this step, and λ is the learning rate. Given a loss function to optimize, the learning rate sets the size of each step used in iteration. The learning rate λ must be selected prior to model training. A fine grid search of values between 0.001 and 1 was executed, as is common in literature [7,13,14]. Based on this grid search, the optimal learning rate was determined to be 0.1, with optimal tree depth of 32, and 500 individual regression trees used to make the final boosted model. The boosted regression tree model was trained and tuned using the training and test sets. Like the random forest model, data collected between June 16th 2020 and July 17th 2020 were never used to adjust the model, and set aside to provide accurate error metrics. The mean average error on this validation data set was 5.38×10^{-15} $[\mathrm{m}^{-\frac{2}{3}}]$, which is slightly higher than the random forest, but compares favorably to all prior explicit models. Although the error reduction between the boosted model and the random forest model was smaller than that between the random forest model and the linear model, it was noticably more appropriate for low-turbulence propogation environments. This trend is visualized in the comparison of measured and predicted $Log_{10}C_n^2$ in Fig. (11).

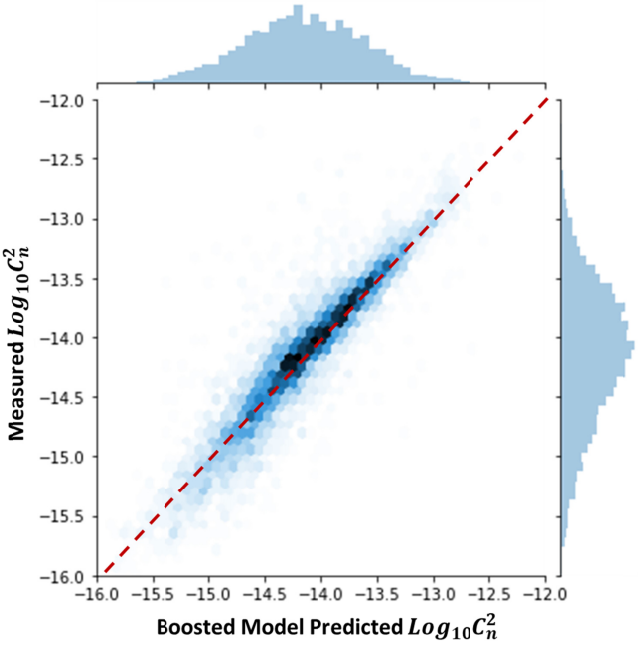


Fig. 11. Joint distribution predicted and measured $Log_{10}C_n^2$ for the boosted model test data between January 1st 2020 and July 17th 2020.

The potential improvement in the boosted model over the random forest model is most evident in the distribution of predicted $Log_{10}C_n^2$ values. Both the measured and predicted distributions appear similarly uniform with a mean just below 1×10^{-14} $[\mathrm{m}^{-\frac{2}{3}}]$, with long tails. Similar to the random forest model, the boosted model displayed reasonable prediction accuracy for high-turbulence environments, even those with measured C_n^2 above 1×10^{-13} $[\mathrm{m}^{-\frac{2}{3}}]$. The bottom-left corner of the joint distribution indicates that the boosted model was effective in predicitng C_n^2 in low turbulence environments, down to around 5×10^{-16} $[\mathrm{m}^{-\frac{2}{3}}]$.

The joint distributions in Fig. (4)-(6), (7), and (10)-(11) provide a concise visualization for the prediction accuracy of each respective model over a long time-frame. Additional insights into model prediction accuracy can be achieved through plotting measured C_n^2 against each model's predicted C_n^2 over the couse of a series of representative days. In order to maintain effective error metrics, only days in the validation set between June 16th 2020 and July 17th 2020 were used in this set, as no model had used these observations for training.

6. MODEL COMPARISON

In order to better compare the relative efficacy of the literature models and new polynomial and machine-learning based models, the predicted C_n^2 for each was compared over the course of four representative days between June 16th 2020 and July 17th 2020. This date range was selected to validate and compare the models, as no model used data collected in this interval for training or hyper-parameter tuning. Thus, no model could make use of a priori knowledge of the relationship between measured C_n^2 and the relevant environmental parameters when making its prediction. Based on the test set performance of the polynomial regression, boosted, and random forest models, the boosted model was omitted from the plots, as it tended to track the random forest model, but with higher error. The model comparison in terms of mean absolute error for C_n^2 prediction is given in Table 3.

Table 3. Environmental Parameters Used for Model Training

Model	Mean Absolute Error
Model 1 (in ref. [8])	1.67×10^{-14} $[m^{-\frac{2}{3}}]$
Model 2 (in ref. [9])	1.61×10^{-14} $[m^{-\frac{2}{3}}]$
Model 3 (in ref. [10])	2.68×10^{-14} $[m^{-\frac{2}{3}}]$
Polynomial Regression	5.47×10^{-15} $[m^{-\frac{2}{3}}]$
Boosted Model	5.35×10^{-15} $[m^{-\frac{2}{3}}]$
Random Forest	5.38×10^{-15} $[m^{-\frac{2}{3}}]$

The four days selected for visual analysis and comparison were June 17th, June 27th, July 10th, and July 13th 2020, as they contained no missing data and had interesting weather features. For each day selected, measurements and predictions were plotted between $12{:}00$ AM and $11{:}59$ PM. The responsiveness of each model to weather changes was analyzed over the course of each day. By analyzing the predicted C_n^2 for each model over the course of one day, the models can be compared visually, and examined for dependence on the measured environmental parameters.

On June 17th 2020, the propagation environment had little wind, and became more humid as the atmospheric pressure decreased over the course of the day. The air temperature was greater than the water temperature for the full day, and clouds limited the magnitude of solar radiation affecting the local atmosphere. The air temperature ranged between about 19 [°C] and 23 [°C], the air-water temperature difference fell between about -2 [°C] and -4 [°C], solar radiation never exceeded 400 $[\frac{\mathrm{W}}{\mathrm{m}^2}]$, pressure fell from 1026 [mBar] to 1020 [mBar], and the wind speed was between 1 $[\frac{\mathrm{m}}{\mathrm{s}}]$ and 3 $[\frac{\mathrm{m}}{\mathrm{s}}]$. The predicted C_n^2 from each newly trained model is presented as a time-series along with the environmental parameters in Fig. (12).

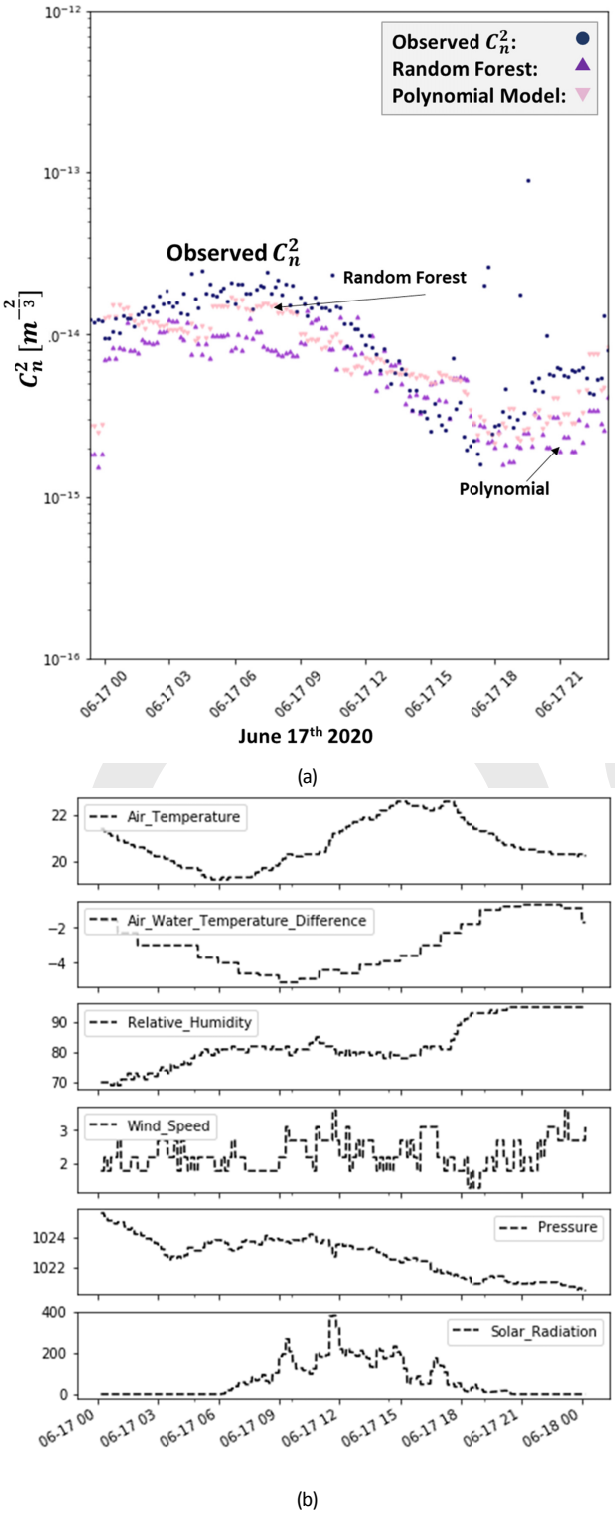
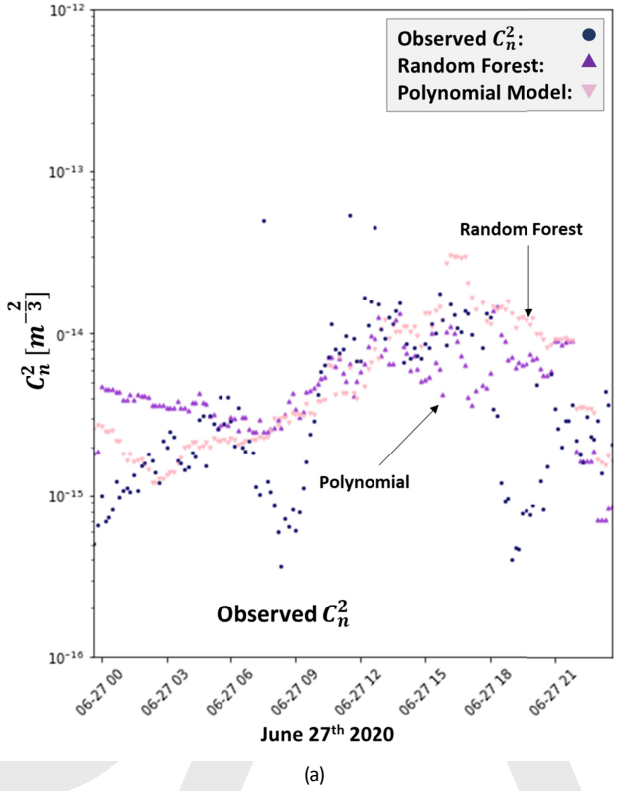
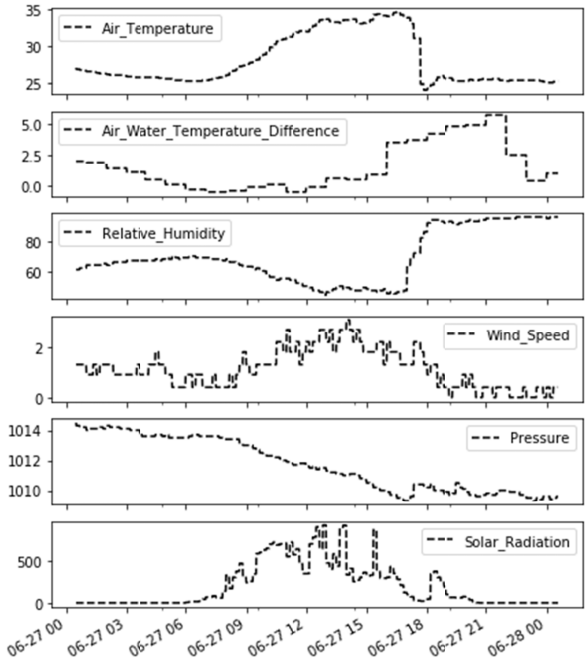


Fig. 12 Measured and predicted C_n^2 for polynomial and random forest models for June 17th 2020 (a), along with measured environmental parameters of interest (b).

Based on the time-series plot of measured C_n^2 and predicted C_n^2 for each of the selected models, both the polynomial and random forest models tracked the measured C_n^2 fairly well.

Next, on June 27th 2020, the propagation environment had little wind, and had below-average humidity. The air temperature was relatively high, and greater than the water temperature for the full day. There were some clouds limiting the magnitude of solar radiation affecting the local atmosphere, but less than on June 16th. The air temperature ranged between about 25 [°C] and 34 [°C], the air-water temperature difference fell between about 0 [°C] and 5 [°C], solar radiation never exceeded 700 [$\frac{W}{m^2}$], pressure fell from 1014 [mBar] to 1010 [mBar], and the wind speed was between 0 $\left[\frac{m}{s}\right]$ and 2 $\left[\frac{m}{s}\right]$. The predicted C_n^2 from each newly trained model is presented as a time-series along with the environmental parameters in Fig. (13).



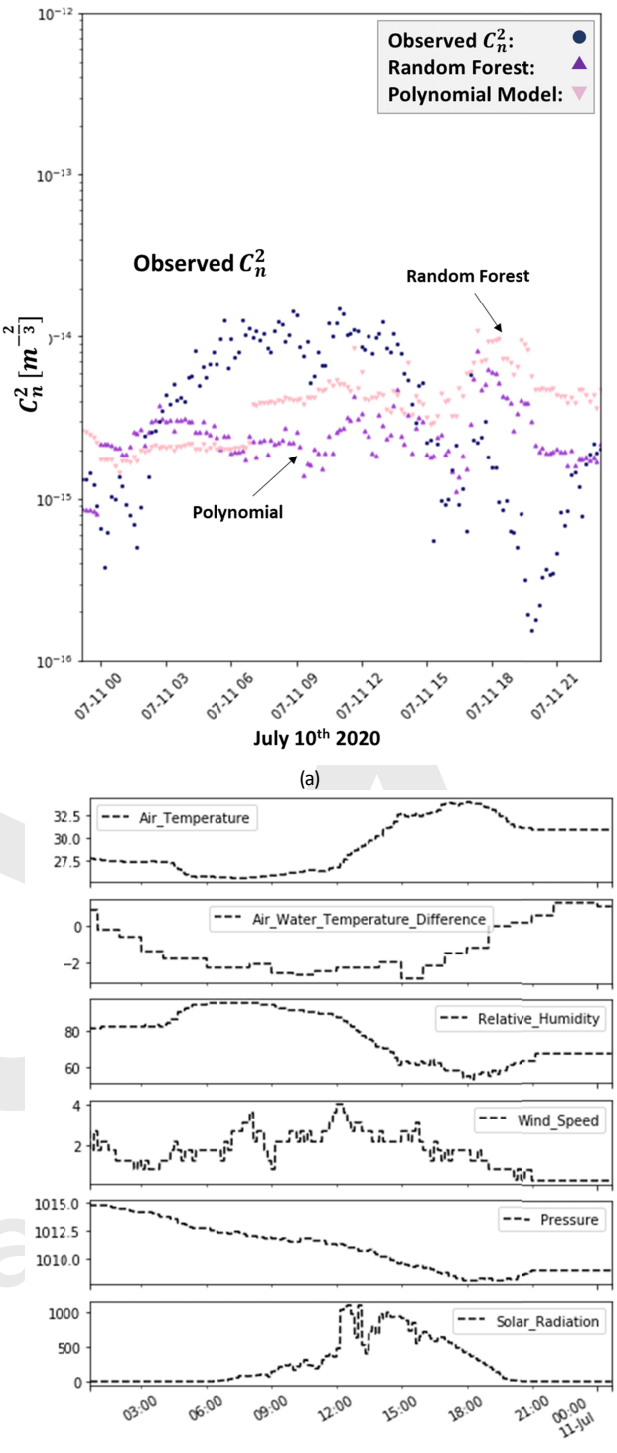


(b)

Fig. 13. Measured and predicted C_n^2 for polynomial and random forest models for June 27th 2020 (a), along with measured environmental parameters of interest (b).

Based on the time-series plot of measured C_n^2 and predicted C_n^2 for each of the models, the polynomial and random forest models tracked changes in C_n^2 relatively worse than on June 16th.

After analyzing the two selected days in June, the same process was applied for July 10th 2020. On that day, the propagation environment had higher than average wind for the first half of the day and became less humid as the atmospheric pressure decreased. The measured air temperature was lower than the water temperature for the first half of the day, and morning clouds affected solar radiation until around noon. The air temperature ranged between about 26 [°C] and 33 [°C], the air-water temperature difference fell between about -2 [°C] and 1 [°C], solar radiation was as high as $1000\,[\frac{\mathrm{W}}{\mathrm{m}^2}]$, pressure fell from $1016\,[\mathrm{mBar}]$ to $1010\,[\mathrm{mBar}]$, and the wind speed was between $1\,\left[\frac{\mathrm{m}}{\mathrm{s}}\right]$ and $3\,\left[\frac{\mathrm{m}}{\mathrm{s}}\right]$. The predicted C_n^2 from each newly trained model is presented as a time-series along with the environmental parameters in Fig. (14).



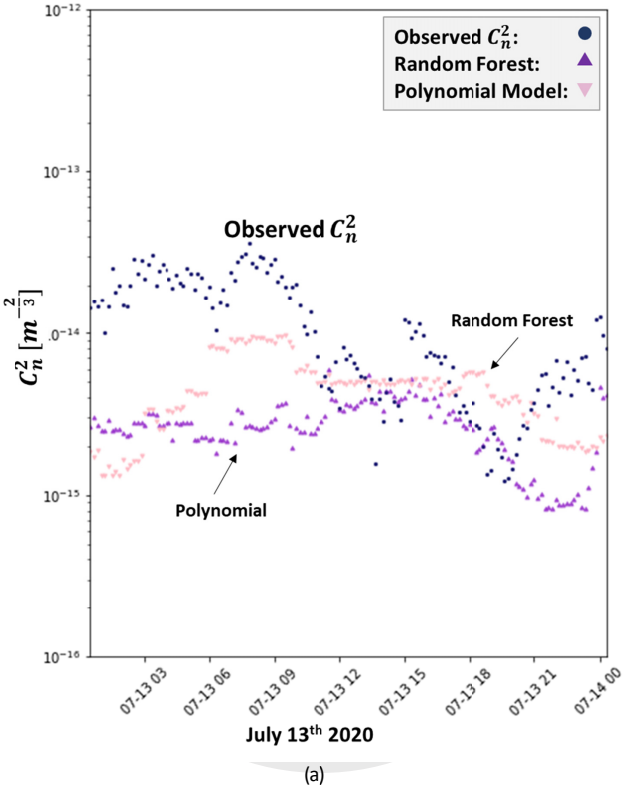
(a)

(b)

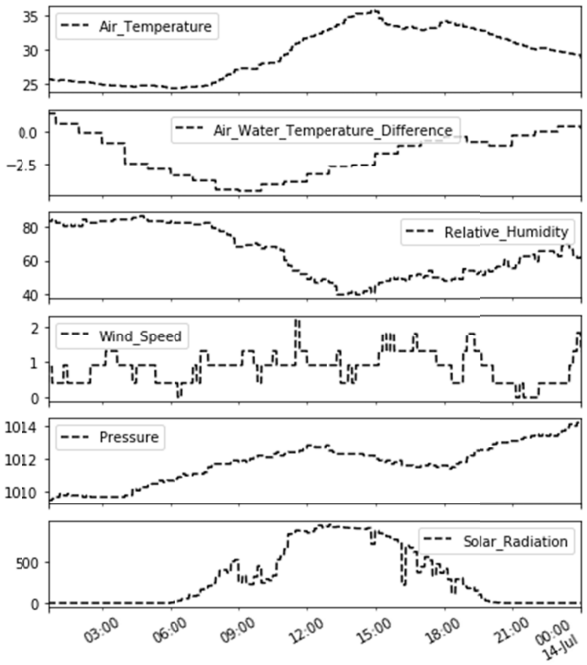
Fig. 14. Measured and predicted C_n^2 for polynomial and random forest models for July 10th 2020 (a), along with measured environmental parameters of interest (b).

Based on the time-series plot of measured C_n^2 and predicted C_n^2, both the polynomial and random forest models tracked the measured C_n^2 fairly well, although they failed to predict the continued drop in optical turbulence during the afternoon.

Finally, the comparison methodology was applied to the data collected on July 13th 2020. The propagation environment had lower than average wind for the full day, and became less humid as the atmospheric pressure increased. The measured air temperature was lower than the water temperature for the first half of the day, and patchy morning clouds affected solar radiation until around 10:00 AM local time. The air temperature ranged between about $25\,[°\mathrm{C}]$ and $36\,[°\mathrm{C}]$, the air-water temperature difference fell between about $-3\,[°\mathrm{C}]$ and $1\,[°\mathrm{C}]$, solar radiation was as high as $1000\,[\frac{\mathrm{W}}{\mathrm{m}^2}]$, pressure rose from $1010\,[\mathrm{mBar}]$ to $1015\,[\mathrm{mBar}]$, and the wind speed was between $0\,\left[\frac{\mathrm{m}}{\mathrm{s}}\right]$ and $2\,\left[\frac{\mathrm{m}}{\mathrm{s}}\right]$. The predicted C_n^2 from each newly trained model is presented as a time-series along with the environmental parameters in Fig. (15).



(a)



(b)

Fig. 15. Measured and predicted C_n^2 for polynomial and random forest models for July 13th 2020 (a), along with measured environmental parameters of interest (b).

On July 13th 2020, the random forest model tracked the measured optical turbulence slightly better than the polynomial model. The relative performance of the random forest and polynomial models on June 16th, June 27th, July 10th, and July 13th suggests that the newly-trained random forest model has some ability to forecast optical turbulence for the one-month period following data collection.

7. CONCLUSION

Using scintillometer data and environmental parameter measurements collected over a near-maritime propagation path between January 1st 2020 and July 17th 2020, six macro-meteorological models were analyzed and compared. Three models from existing literature were analyzed over the full time period. They were found to generally over-predict the extent of optical turbulence. Models 1 and 2 both tended to predict C_n^2 in a tight cluster between $7 \times 10^{-15}[\mathrm{m}^{-\frac{2}{3}}]$ and $1 \times 10^{-13}[\mathrm{m}^{-\frac{2}{3}}]$, and tended not to change considerably over the course of each day. In contrast, Model 3 tended to over-predict the extent of optical turbulence to a greater extent but was able to track some changes in C_n^2 over the course of the day. This suggests that Model 3 could be adjusted using a scaling factor to better predict the extent of optical turbulence in a near-maritime environment.

The polynomial regression, boosted regression tree, and random forest models tended to track changes in C_n^2 better than the literature models, but did not always generate accurate predictions. Of these models, the polynomial model was the simplest in terms of the number of parameters, and the random forest model had the lowest mean average error. This could potentially be partially attributed to tree over-fitting and mitigated through further hyper-parameter tuning. The fact that the error in boosted regression tree predictions was generally higher than random forest prediction error suggests that further hyper-parameter tuning could improve prediction accuracy.

Comparing the relative efficacy of these six models suggests that machine-learning techniques can be used to train improved models for predicting C_n^2 in the near-maritime environment. The stability of these models allowed for some forecasting ability, which suggests that models were not over-fit to training data.

Further investigation into the relationships between environmental parameters and C_n^2 in maritime environments will allow for higher prediction accuracy for new machine-learning informed models. Although the six environmental parameters selected for training were suggested by prior literature to be appropriate [7], the scope of appropriate predictor variables will be investigated further. Additionally, applying these new models to other propagation environments will yield insight into their wider applicability. Studying the effect of propagation environment on prediction accuracy will be important in determining where and how these models can be successfully applied.

A wider variety of machine-learning based model architectures will be applied to the data set, in order to further improve prediction accuracy in the near-maritime environment. Additionally, by expanding the duration of the data set, the temporal and seasonal dependencies in machine-learning model training and prediction accuracy could be investigated. These models will be investigated further once a full year of data is available.

Funding Information. This work is supported by The Office of Naval Research, the Directed Energy Joint Technology Office, and the United States Naval Academy Trident Scholar Program.

Acknowledgement. This work could not have been accomplished without instruction and review from Professors William Traves, Douglas Vanderwerken, Daren Creutz, Daniel Roche, Emily Retzlaff, and CAPT Chris Ruth, USN. The authors would also like to thank the United States Naval Academy Mechanical Engineering, Electrical Engineering, Oceanography and Mathematics Departments for their assistance in data collection, model critique, and editorial review. Finally, we would like to thank the Trident Committee and Professor Maria Schroeder for their generous support.

Disclosures. The authors declare no conflicts of interest.

References

1. Ricardo Barrios and Federico Dios (March 7th 2012). Wireless Optical Communications Through the Turbulent Atmosphere: A Review, Optical Communications Systems, Narottam Das, IntechOpen, DOI: 10.5772/34740.
2. Andrews, L. C., & Phillips, R. L. (2005). Laser beam propagation through random media. SPIE.
3. Frederickson, P. A., Hammel, S., & Tsintikidis, D. (2006, September). Measurements and modeling of optical turbulence in a maritime environment. In Atmospheric Optical Modeling, Measurement, and Simulation II (Vol. 6303, p. 630307). International Society for Optics and Photonics.
4. Mahon, R., Moore, C. I., Burris, H. R., Rabinovich, W. S., Stell, M., & Thomas, L. M. (2009). Analysis of long-term measurements of laser propagation over the Chesapeake Bay. Applied optics, 48(12), 2388-2400.
5. Frederickson, P. A., Davidson, K. L., Zeisse, C. R., & Bendall, C. S. (2000). Estimating the refractive index structure parameter over the ocean using bulk methods. Journal of Applied Meteorology, 39(10), 1770-1783.
6. Bourque W., Nelson C., Nelson D, "Evidence and Implications of Differences in Atmospheric Optical Turbulence Behavior on Opposite Coastal Environments," Journal of Directed Energy 6, Spring 2017, 187-197.
7. Jellen, C., Burkhardt, J., Brownell, C., & Nelson, C. (2020). Machine learning informed predictor importance measures of environmental parameters in maritime optical turbulence. Applied Optics, 59(21), 6379-6389.
8. Sadot, D., & Kopeika, N. S. (1992). Forecasting optical turbulence strength on the basis of macroscale meteorology and aerosols: models and validation. Optical Engineering, 31(2), 200-213.
9. Wang, H., Li, B., Wu, X., Liu, C., Hu, Z., & Xu, P. (2015). Prediction model of atmospheric refractive index structure parameter in coastal area. Journal of Modern Optics, 62(16), 1336-1346.
10. Raj, A. A. B., Selvi, J. A. V., & Durairaj, S. (2015). Comparison of different models for ground-level atmospheric turbulence strength (Cn 2) prediction with a new model according to local weather data for FSO applications. Applied optics, 54(4), 802-815.
11. Oermann, R.J. Novel Methods for the Quantification of Atmospheric Turbulence Strength in the Atmospheric Surface Layer. Ph.D. Thesis, School of Chemistry and Physics, University of Adelaide, Adelaide, SA, Australia, 2014. Available at digital.library.adelaide.edu.au.
12. Jellen, C., Nelson, C., Brownell, C., Burkhardt, J., & Oakley, M. (2020). Measurement and analysis of atmospheric optical turbulence in a near-maritime environment. IOP SciNotes, 1(2), 024006.
13. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.
14. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112, pp. 3-7). New York: springer.
15. Strobl, C., Boulesteix, A. L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC bioinformatics*, *8*(1).
16. US National Oceanic and Atmospheric Administration. Annapolis, MD Tides and Currents. Retrieved February 10, 2021 from https://tidesandcurrents.noaa.gov/stationhome.html?id=8575512
17. Scintec AG. BLS450 Boundary Later Scintillometer Technical Specifications. Retrieved November 22, 2020 from https://scintec.com/english/CustomUpload/37403570340037003560369035003400365036003730356036903700372036402/BLS450.pdf
18. Google Maps, 2020. United States Naval Academy, 1:500. Google Maps [online] Available at: https://www.google.com/maps/@38.9833615,-76.4738069,1422m/data=!3m1!1e3 [Accessed 30 November 2020].
19. Davis Instruments Corporation. Vantage Pro 2 Wireless Sensor Suite and Technical Specifications. Retrieved November 22, 2020 from https://www.davisinstruments.com/product_documents/weather/spec_sheets/6152_62_53_63_SS.pdf
20. US Department of Commerce, National Oceanic and Atmospheric Administration, National Weather Service, & National Data Buoy Center. (1996, November 8). National Data Buoy Center. Retrieved from https://www.ndbc.noaa.gov/measdes.shtm
21. NOAA ESRL Improved Sunrise/Sunset Calculator. (n.d.). Retrieved April 17, 2020, from https://www.esrl.noaa.gov/gmd/grad/solcalc/sunrise.html
22. Jellen, C., Atmospheric Research Repository, (2020), GitHub repository, https://github.com/CDJellen/DataRepository