



Effective Benchmarks for Optical Turbulence Modeling

Christopher Jellen,^a Charles Nelson,^b Cody Brownell,^a and John Burkhardt^a

^a *United States Naval Academy Mechanical Engineering Department*

^b *United States Naval Academy Electrical and Computer Engineering Department*

Corresponding author: Christopher Jellen, cdjellen@gmail.com

Early Online Release: This preliminary version has been accepted for publication in *Artificial Intelligence for the Earth Systems*, may be fully cited, and has been assigned DOI 10.1175/AIES-D-24-0003.1. The final typeset copyedited article will replace the EOR at the above DOI when it is published.

© 2024 American Meteorological Society. This is an Author Accepted Manuscript distributed under the terms of the default AMS reuse license. For information regarding reuse and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

ABSTRACT: Optical turbulence poses a significant challenge for communication, directed energy, and imaging systems, particularly in the atmospheric boundary layer. Effective modeling of optical turbulence is crucial for the development and deployment of these systems, yet the lack of standardized evaluation tools and benchmark data sets hinders the development and adoption of machine learning to address these challenges. We introduce the `otbench` Python package, a comprehensive framework for rigorous development and evaluation of optical turbulence strength prediction models. This package provides a consistent interface for testing models across diverse data sets and tasks, including a novel, long-term data set collected over two years at the United States Naval Academy. `otbench` incorporates a range of baseline models (statistical, data-driven, and deep learning), enabling researchers to assess the relative quality of their approaches and identify areas for improvement. Our analysis reveals the applicability of various models across different environments, highlighting the importance of long-term data sets for robust model evaluation. By promoting standardized benchmarking and facilitating model comparison, `otbench` empowers researchers to accelerate the adoption of machine learning techniques for optical turbulence modeling.

SIGNIFICANCE STATEMENT: The accurate prediction and forecasting of optical turbulence is vital for the successful operation of laser-based systems in operational environments. However, the lack of standardized benchmark data sets and evaluation tools has hindered progress in this field, particularly in the application of advanced machine learning techniques. The `otbench` Python package addresses this critical gap by providing a comprehensive framework for developing, evaluating, and comparing optical turbulence models. By incorporating a novel, long-term data set and a range of baseline models, `otbench` enables rigorous model assessment and facilitates the development of next-generation machine learning solutions that can significantly enhance the performance and reliability of laser-based systems in real-world conditions. This package not only contributes to the scientific understanding of optical turbulence but also holds significant potential to drive innovation in the design and optimization of machine learning applications in laser-based technologies across various domains, including communication, directed energy, and imaging.

1. Introduction

Optical turbulence is a major consideration in the design and deployment of communication, directed energy, and imaging systems (Andrews and Phillips 2005). These systems must operate effectively in diverse propagation environments, across a wide range of turbulent intensities. Optical turbulence effects are quantified using the refractive index structure parameter, C_n^2 (Andrews and Phillips 2005). This parameter communicates the magnitude of fluctuations in the refractive index. Higher values of C_n^2 indicate stronger optical turbulence, which increases the deleterious impact of effects such as beam wander and spreading on a wavefront as it travels from a transmitter to a receiver (Andrews and Phillips 2005). The accurate modeling of optical turbulence strength in these environments is thus a key requirement for ensuring the performance and reliability of these critical systems. However, modeling optical turbulence in the boundary layer from bulk-atmospheric measurements is a complex and challenging problem, leading to the development of numerous models (Barnett et al. 2022; Frederickson et al. 2000; Mahon et al. 2020; Sadot and Kopeika 1992; Wang et al. 2015; Wang and Basu 2016; Jellen et al. 2020a, 2023; Vorontsov et al. 2020).

Historically, the lack of standardized benchmark data sets and evaluation tools has limited the ability to rigorously compare and assess the performance of different optical turbulence models.

This has also impeded the application of more advanced machine learning (ML) techniques, which often rely on large, diverse data sets for model development or training. The importance of benchmark data sets in driving ML innovation is evident in fields like weather forecasting, where frameworks including `WeatherBench` were instrumental in identifying the utility of residual connections and fully-convolutional models in generating effective forecasts (Rasp et al. 2020).

To address these challenges and foster the development of novel solutions, we introduce the `otbench` Python package. This package aims to provide a comprehensive framework for developing, evaluating, and comparing optical turbulence strength prediction models. By offering a standardized interface for accessing and processing various data sets, defining benchmark tasks, and evaluating model performance, `otbench` establishes a solid foundation for researchers to rigorously compare existing models and develop new, potentially ML-based, approaches. Baseline statistical, data-driven, and deep-learning model implementations illustrate the significant opportunity to develop improved, sophisticated ML models which generalize across a range of propagation environments.

A key highlight of `otbench` is the inclusion of the first publicly available long-term measurement campaign, collected over two years using the United States Naval Academy scintillometer link. This unique data set, coupled with standardized evaluation metrics, empowers researchers to assess model performance across a wide range of temporal and environmental conditions, offering a critical benchmark for developing models that can generalize well to real-world scenarios. These features, and the ML applications they enable, will ultimately enhance the performance of laser-based systems in diverse operational environments.

2. Overview of `otbench`

Effective performance benchmarks require fixed objectives, evaluation metrics, and data sets. In the context of data-driven and deep learning models, this also includes fixed and transparent training, testing, and validation splits for each data set. The `otbench` package brings these standards to the challenge of optical turbulence modeling. The implementation is inspired by packages such as `WeatherBench` developed by Rasp et al. (2020) and the Open Graph Benchmark `ogb` developed by Hu et al. (2020), which provide researchers with a foundation for weather forecasting and for graph learning respectively. By demonstrating the applicability of standard deep-learning and

other ML models to the domain of optical turbulence modeling, `otbench` will demonstrate the opportunity associated with developing more advanced ML models and build trust with the existing community of researchers and practitioners.

a. Common tasks for modeling turbulence strength

A range of models exist to predict the strength of boundary layer optical turbulence from meteorological data (Sadot and Kopeika 1992; Wang et al. 2015; Wang and Basu 2016; Pierzyna et al. 2023; Frederickson et al. 2000; Jellen et al. 2023). Establishing comparisons between these models requires careful re-implementation by each successive researcher. The lack of accessible field data further limits reproducibility. Open data, explicit model implementations, and reproducible predictions are essential in evaluating new models in context, and in making comparisons against prior approaches.

The `otbench` package is built around *tasks*, under which data sets are processed and standardized to enable comparisons across approaches. Data sets are developed from real-world field experiments and include a range of meteorological and oceanographic measurements. Research teams currently decide how to perform quality control on the data captured from these field experiments. Common approaches include resampling to a standard frequency, interpolating to fill missing measurements, or dropping records in which one or more measurement is missing. While each approach is valid, the lack of standards prevents effective comparisons between the models these teams develop. The `otbench` package uses tasks to enforce standard processing of data sets to mitigate this challenge. Tasks are explicit in their handling of missing measurements, data loading, and feature availability. Each task also references the model's target (measured C_n^2 at a given height of observation) and the evaluation metrics under which a model's performance is measured. This enables effective comparison between models, and understanding of a model's relative performance against prior literature models.

b. Domain diversity

The absence of standard benchmark tasks and data sets, especially long-term data sets, contribute to the challenge of over-fitting to local micro-climates. The `otbench` package attempts to address both of these concerns. The package currently includes data from both the Mona Loa C_n^2 Study

described in Maclean and Oncley (2006) and the United States Naval Academy Long Term Scintillation Study (Jellen et al. 2024). Both data sets measure C_n^2 in the atmospheric boundary layer alongside local meteorological data. Modeling approaches which perform well on both data sets may be robust to a wider range of propagation environments.

The United States Naval Academy Long Term Scintillation Study includes over two years of scintillometer data along with a range of meteorological and oceanography parameters (Jellen et al. 2021). This temporal range ensures models are evaluated across a range of seasonal conditions. In the context of ML and data-driven models, it also enables deeper investigation into the relationship between time and prediction error.

c. Using otbench

The `otbench` package seeks to lower barriers to entry in evaluating existing models and developing new models for optical turbulence strength prediction. In addition to supplying benchmark tasks and data sets, `otbench` includes utilities for defining new models and evaluating their performance against a given task. Data-driven and deep learning models may be implemented as sub-classes of the base regression or forecasting model, offering data-processing and evaluation utilities provided by `otbench`. Bulk-atmospheric models with explicit parameters are also easy to develop under this context. The package includes sample implementations of each type of model, further described in Section 5. By adopting `otbench`, researchers are empowered to focus on crafting new models, with the package defining the data processing, model evaluation, and providing performance metrics for past implementations. This development flow is described visually in Figure 1.

In Figure 1, the “Task” and “Evaluation” phases are defined by the `otbench` package, while the “Model” phase is defined by the researcher, optionally using the package-defined “Base model” for bulk-atmospheric, data-driven, or deep learning approaches. This base model supplies standard utilities to increase the speed of model development. For deep learning models, researchers need only specify a model architecture and `train` method; the base model supplies data loading functionality. Physics-based approaches which do not rely on local data for training can be defined as a “final model” without the need for other user-defined components. The “Evaluation” phase can evaluate model performance in both cases, optionally storing performance results for contextualization and comparison. For all user-defined models, the “Evaluation” is performed

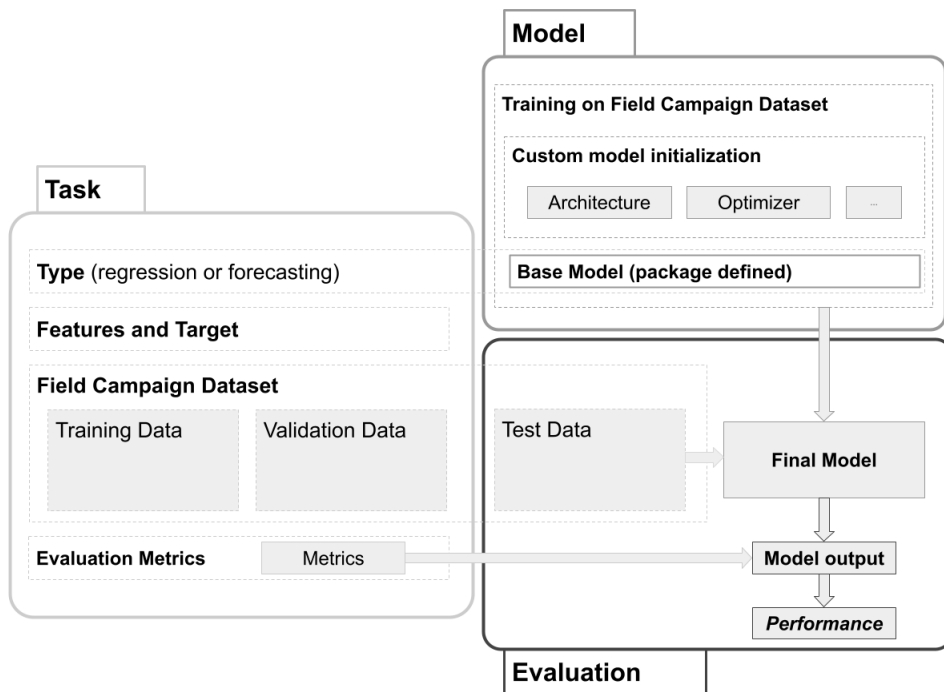


FIG. 1. Model development and benchmark workflow under the otbench package.

against the task’s test data. This imposes a requirement that the user-defined model generate predictions without the use of atmospheric data not measured during the underlying field study.

3. Data

a. Mona Loa Observatory (MLO) C_n^2 Study

Optical turbulence is a key concern for earth-based observatories. The High Altitude Observatory (HAO) manages a solar observatory at Mona Loa, Hawai’i (Maclean and Oncley 2006). The HAO and the National Center for Atmospheric Research (NCAR) conducted a measurement campaign to study optical turbulence in the boundary layer during the summer of 2006 with the support of the Integrated Surface Flux Facility (ISFF) (NCAR/EOL ISFS Team 2023). This study used sonic anemometers fixed at a series of heights to quantify the refractive index structure parameter C_n^2 , along with meteorological features. Instruments were deployed on an existing 40m tower operated by the National Oceanographic and Atmospheric Administration (NOAA) at the Mauna

TABLE 1. Overview of the MLO C_n^2 data set.

Long Name	Short Name	Description	Units	Heights [m]
Raw voltage	kh2oV	Raw voltage	V	6, 13, 23
Water vapor density	kh2o	Water vapor density	g m^{-3}	6, 13, 23
U component	u	U component of wind speed	m s^{-1}	6, 13, 23
V component	v	V component of wind speed	m s^{-1}	6, 13, 23
W component	w	W component of wind speed	m s^{-1}	6, 13, 23
Virtual temperature	tc	Virtual temperature from the speed of sound	$^{\circ}\text{C}$	6, 13, 23
Wind direction	Dir	Mean wind direction	$^{\circ}\text{C}$	10
Wind speed	Spd	Mean wind speed	m s^{-1}	10
Air pressure	P	Barometric pressure	mbar	2
Air temperature	T	Ambient air temperature	$^{\circ}\text{C}$	2
Relative humidity	RH	Relative humidity	%	2
Dew point temperature	Tdew	Dew point temperature	$^{\circ}\text{C}$	2
Refractive index structure parameter	C_n^2	Optical turbulence strength	$\text{m}^{-\frac{2}{3}}$	6, 13, 23
Coordinates				
Time	time	The local time at which a measurement was taken	s	-

Loa Observatory. Sonic anemometers were fixed on the east side of the tower at heights of 6m, 13m, and 23m from the base of the tower (NCAR/EOL ISFS Team 2023). The sonic's data was supplemented by additional NOAA data which measured atmospheric features near the surface, with wind data near 10m. Data from this measurement campaign is made available through the University Corporation for Atmospheric Research (UCAR) at <https://data.eol.ucar.edu/dataset/160.007> as a NetCDF4 file.

Measurements are averaged at a 5 minute frequency, and were collected between June 9th, 2006 and August 8th, 2006. An overview of the available data, including the heights of measurement on the observation tower, are available in Table 1. 9,857 of the 14,038 total rows contain no missing measurements.

b. Unites States Naval Academy long-term scintillation study

The Unites States Naval Academy (USNA) long term scintillation study is a continuing effort to characterize and measure optical turbulence in the near-maritime boundary layer. A range of experiments have been conducted across the Severn River in Annapolis, Maryland (Jellen et al. 2021). The longest running of these experiments used a ScinTec BLS450 scintillometer to characterize the strength of optical turbulence over a link approximately 3 m above the river’s surface between January 1st, 2020, and July 26th, 2023 (Jellen et al. 2023). This field campaign included a range of meteorological and oceanographic measurements from nearby NOAA, National Data Buoy Center (NDBC), and Academy weather stations.

The two subsets of this campaign are used as data sets. The first, denoted `usna_cn2_sm`, includes 3 months in the summer of 2021 in which few measurements were missing from the scintillometer data set, supplemented with NOAA Coastal Observation data describing the local atmosphere and water conditions (Center for Operational Oceanographic Products and Services 2023). The second covered the period between January 1st, 2020 and July 14th, 2022 and is denoted `usna_cn2_lg`. This data set includes local measurements made with a Davis Vantage Pro2 weather station, as well as water conditions from the NDBC Thomas Point data buoy (Jellen et al. 2020b; National Data Buoy Center 2023). Future releases will develop a full 42 month data set, including measurements from all sources as available.

1) USNA C_n^2 SHORT-DURATION DATA SET

Field campaigns studying optical turbulence present a significant expense for researchers. Some capture measurements over a small number of days or weeks using off-the-self weather stations rather than high-frequency sonics (Sadot and Kopeika 1992; Wang et al. 2015). The low cost and availability of these weather stations may present an opportunity to develop models which perform across disparate propagation environments.

The `usna_cn2_sm` data set includes measured C_n^2 along with atmospheric and oceanographic parameters from a co-located NOAA observation station between June 1st, 2021, and September 1st, 2021. These summer months were selected to roughly map to the time-of-year used in the MLO C_n^2 study, and due to the low number of missing measurements from the scintillometer. The `usna_cn2_sm` data set includes parameters at a single level of observation, with a 6-minute

TABLE 2. Overview of the USNA C_n^2 Small data set.

Long Name	Short Name	Description	Units	Heights [m]
Wind direction	Dir	Mean wind direction	m s^{-1}	10
Wind speed	Spd	Mean wind speed	m s^{-1}	10
Air pressure	P	Barometric pressure	mbar	10
Air temperature	T	Ambient air temperature	$^{\circ}\text{C}$	5
Relative humidity	RH	Ambient relative humidity	%	2
Water temperature	T	Water surface temperature	$^{\circ}\text{C}$	0
Solar radiation	Rad	Incoming total solar radiation at ground level	$\text{W}^2 \text{m}^{-1}$	1
Refractive index structure parameter	C_n^2	Optical turbulence strength	$\text{m}^{-\frac{2}{3}}$	3
Coordinates				
Time	time	The local time at which a measurement was taken	s	-
Latitude	lat	The latitude at which the scintillometer and weather station are located	$^{\circ}$	-
Longitude	lon	The longitude at which the scintillometer and weather station are located	$^{\circ}$	-
Altitude	alt	The height above sea level of the scintillometer	m	-

averaged frequency of observation. A detailed description is presented in Table 2. 21,007 of the 22,081 total observations have no missing measurements. As a result, no interpolation is applied to the data set. The usna_cn2_sm data set captures the rapid fluctuations in C_n^2 , as well as the weaker diurnal cycle for optical turbulence in the near-maritime environment above the Severn River.

2) USNA C_n^2 LONG-DURATION DATA SET

In contrast to the usna_cn2_sm data set, the usna_cn2_lg data set includes data measured at or interpolated to a 1-minute frequency, with some significant gaps due to instrument misalignment or power outages. Interpolation is applied within reporting intervals, but not across gaps where measurements are missing. This data set offers the first known opportunity to develop models for predicting or forecasting optical turbulence over more than two years. The periods of missing

TABLE 3. Overview of the USNA C_n^2 Large data set.

Long Name	Short Name	Description	Units	Heights [m]
Wind direction	Dir	Mean wind direction	m s^{-1}	3
Wind speed	Spd	Mean wind speed	m s^{-1}	3
Air pressure	P	Barometric pressure	mbar	3
Air temperature	T	Ambient air temperature	$^{\circ}\text{C}$	3
Relative humidity	RH	Ambient relative humidity	%	3
Water temperature	T	Water surface temperature	$^{\circ}\text{C}$	0
Solar radiation	Rad	Incoming total solar radiation at ground level	$\text{W}^2 \text{m}^{-1}$	1
Temporal hour	th	The time elapsed since sunrise divided by 1/12 the time between sunset and sunrise for a given day	-	-
Refractive index structure parameter	C_n^2	Optical turbulence strength	$\text{m}^{-\frac{2}{3}}$	3
Coordinates				
Time	time	The local time at which a measurement was taken	s	-
Latitude	lat	The latitude at which the scintillometer and weather station are located	$^{\circ}$	-
Longitude	lon	The longitude at which the scintillometer and weather station are located	$^{\circ}$	-
Altitude	alt	The height above sea level of the scintillometer	m	-

measurements present an additional challenge for model development, better reflecting real-world conditions. A detailed description is presented in Table 3. 1,155,041 of the 1,291,225 total observations have no missing measurements after interpolation to re-index to a 1-minute frequency. This data set presents an opportunity to study model performance across seasons and over longer time scales. It further presents an opportunity to assess a model over a validation set of one year or longer, while training with data across all seasons.

c. Adding new field campaigns

As currently implemented, `otbench` includes three data sets from two field campaigns. While this provides researchers with a measure of environmental extensibility for their models, new data sets captured in different propagation environments will increase validation quality. The package currently stores data sets under the NetCDF4 format, using time as a required coordinate. There are no specific requirements for the instruments used to capture data or handling of missing measurements. A standard frequency of observation is assumed when developing and evaluating forecasting models; however, the package does not impose strict requirements on formatting or metadata requirements on its constituent data sets. New data sets are best contributed directly to the source repository at <https://github.com/cdjellen/otbench>.

4. Benchmark Tasks

While standard data sets are necessary for comparing model performance, especially across the literature, they are not sufficient for robust evaluation. Fixed training and validation sets for data-driven modeling approaches, a fixed test set for all models, standard evaluation metrics, and standard handling of missing measurements improve the quality and robustness of benchmark evaluation. The `otbench` package includes regression tasks and forecasting tasks. Under regression tasks, models predict the strength of optical turbulence from a set of meteorological measurements. In forecasting tasks, the measured C_n^2 some fixed number of observations in the future is predicted from past meteorological and C_n^2 measurements.

All tasks have common structure, specifying a data set, metadata, treatment of missing measurements (either ‘full’ for the complete data set or ‘dropna’ for the data set without missing measurements). These observations are removed from the train, test, and validation sets independently. Each task further specifies the target parameter, typically the name of the C_n^2 column for that data set (such as ‘Cn2_3m’ or ‘Cn2_15m’), and the columns assumed unavailable for training purposes, typically C_n^2 measured at other elevations. Some models which make use of solar features such as the temporal hour require knowledge of the location and timezone of the data set; these are specified at the task level. Finally, tasks specify the train, test, and validation sets based on the index of the transformed data set, whether to take the base-10 log of C_n^2 , and which metrics to evaluate model performance under.

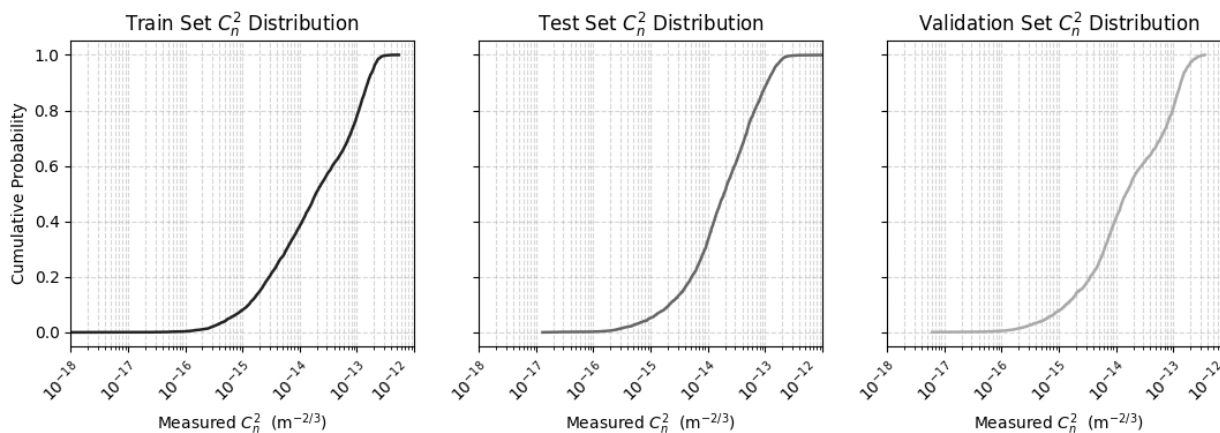


FIG. 2. Cumulative distribution of measured C_n^2 in the MLO data set.

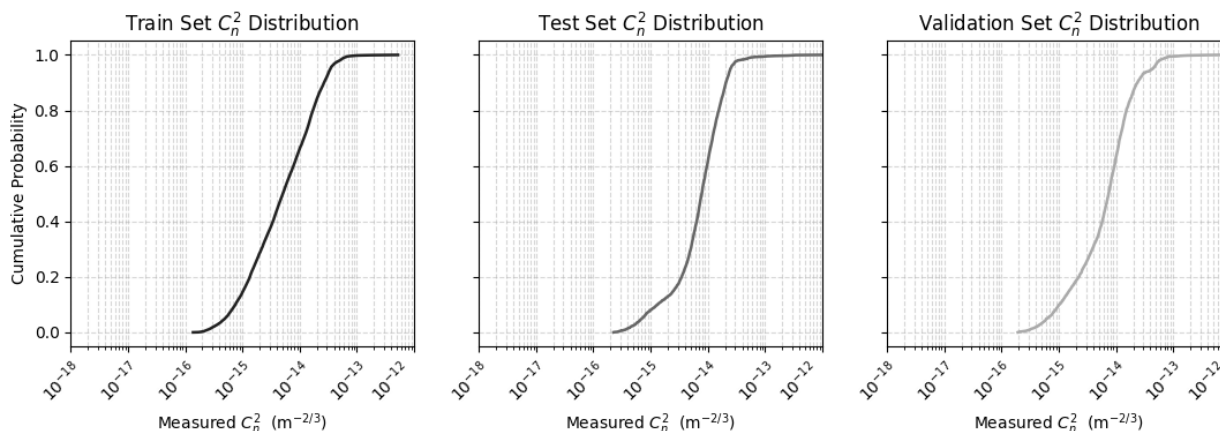


FIG. 3. Cumulative distribution of measured C_n^2 in the USNA short-duration data set.

Both the regression and forecasting tasks order the train, test, and validation indices sequentially. For robust evaluation in operational contexts, it is critical that models perform across a range of seasonal conditions. For regression tasks specifically, C_n^2 measurements are unavailable after a field campaign concludes.

The cumulative distributions of measured C_n^2 in Figure 2, Figure 3, and Figure 4 demonstrate that across all field campaigns few measurements fell below $10^{-15} \text{ m}^{-2/3}$ or above $10^{-12} \text{ m}^{-2/3}$. Within each data set, the train, test, and validation sets do not appear to diverge dramatically. This is especially evident in Figure 4 despite the presence of missing measurements described in Section 2.

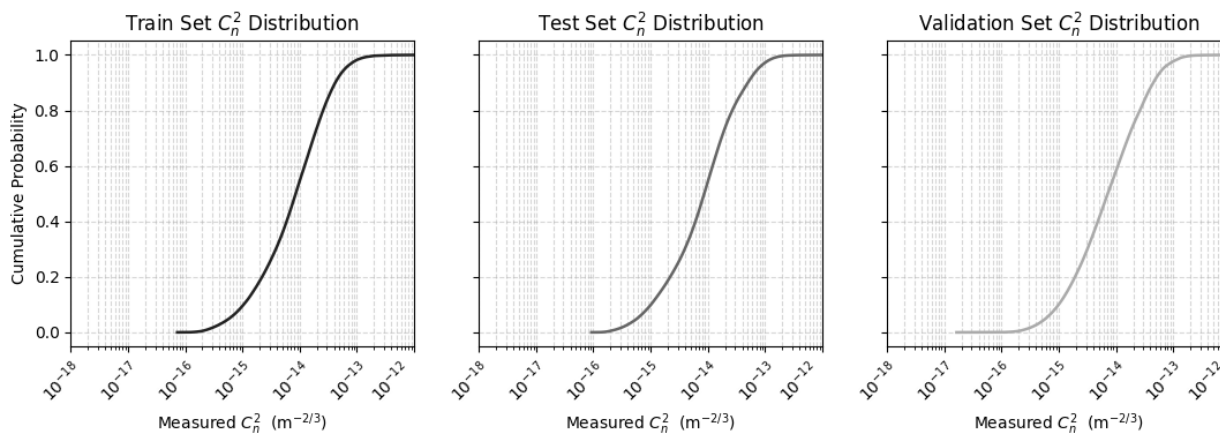


FIG. 4. Cumulative distribution of measured C_n^2 in the USNA long-duration data set.

a. Regression

Regression tasks use a set of standard metrics including the Root Mean Square Error (RMSE), Coefficient of Determination (R^2), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). These metrics were selected in an effort to enable comparisons against prior literature (Pierzyna et al. 2023). For a given regression task, all models are compared across these metrics and the number of valid (non-null) predictions generated by the model. This context offers researchers insight into the relative performance of existing models, as well as the potential improvement afforded by the new models which their teams develop. The current regression tasks are further described in Table 4.

b. Forecasting

Forecasting tasks differ from regression tasks by including a fixed inference window size m and forecast horizon n . The window size specifies the number of past observations available at inference time to predict the strength of optical turbulence n observations in the future. This window includes the past m observations of both meteorological parameters and C_n^2 , in contrast to regression tasks which do not include C_n^2 at inference time.

Forecasting tasks assume direct forecasting, in which the extent of C_n^2 n observations in the future is predicted from available data. This is in contrast to iterative forecasting, in which C_n^2 is predicted 1 observation in the future, and then fed back into the model until the predicted extent of C_n^2 at

TABLE 4. Current regression tasks implemented for otbench.

	MLO C_n^2 (‘full’)	MLO C_n^2 (‘dropna’)	USNA C_n^2 Small	USNA C_n^2 Large
Data set name	mlo_cn2	mlo_cn2	usna_cn2_sm	usna_cn2_lg
Latitude	19.53	19.53	38.98	38.98
Longitude	-155.57	-155.57	-176.48	-176.48
Time zone	US/Hawaii	US/Hawaii	US/Eastern	US/Eastern
Train Indices	[0, 8366]	[0, 8366]	[0, 14639]	[0, 524339]
Test Indices	[8367, 10366]	[8367, 10366]	[14640, 17999]	[524340, 788208]
Validation Indices	[10367, 13942]	[10367, 13942]	[18000, 22080]	[788209, 1291224]
Drop missing	false	true	false	false
log 10 transform	true	true	true	true
Target	C_n^2 15 m	C_n^2 15 m	C_n^2 3 m	C_n^2 3 m

TABLE 5. Current forecasting tasks implemented for otbench.

	MLO C_n^2 (‘dropna’)	USNA C_n^2 Small
Data set name	mlo_cn2	usna_cn2_sm
Latitude	19.53	38.98
Longitude	-155.57	-176.48
Time zone	US/Hawaii	US/Eastern
Train Indices	[0, 8366]	[0, 14639]
Test Indices	[8367, 10366]	[14640, 17999]
Validation Indices	[10367, 13942]	[18000, 22080]
Drop missing	true	true
Log transform	true	true
Target	C_n^2 15 m	C_n^2 3 m

n is available. The current forecasting tasks are further described in Table 5. Forecasting tasks implement the same evaluation metrics as regression tasks described in Section a.

5. Baseline models

The package includes a range of literature and data-driven models as baselines for assessing the relative quality of new models. These models fall under three broad categories; models inspired

by weather forecasting including climatology and persistence as in Rasp et al. (2020), prior macro-meteorological models developed from in-situ measurements as in Sadot and Kopeika (1992) or Wang et al. (2015), and simple deep learning models (Jellen et al. 2023).

Regression models and forecasting models are implemented separately, enabling forecasting models to leverage prior observations of the target to inform predictions at the forecast horizon. These prior measurements are typically unavailable for operational regression models, which predict optical turbulence strength from other measurable parameters. The set of baseline models can be extended to cover more of the literature and novel deep learning approaches to the optical turbulence regression and forecasting tasks. Although not required for future tasks, all current tasks transform measured C_n^2 to the corresponding $\log_{10} C_n^2$.

a. Statistical models

Baseline models provide the context required to characterize the performance of more complex models. The WeatherBench package implements persistence and climatology approaches for the task of weather forecasting (Rasp et al. 2020). These inspire the persistence and climatology models in otbench. In the context of forecasting, these baselines are augmented by simple linear forecasting and mean-window models. The linear forecasting model uses least-squares regression to fit a curve to the window of data available during inference, while the mean window model predicts the mean value of C_n^2 observed during this window.

1) PERSISTENCE

Persistence models serve as a common baseline in weather forecasting contexts (Rasp et al. 2020). These models predict a static value for the target, in this case C_n^2 , for all future observations. In the context of optical turbulence modeling, persistence models take the most recent observation of C_n^2 as the prediction for the next observation. Other models are expected to outperform persistence models, as they do not leverage any information about the macro-meteorological conditions at the time of prediction. Given the diversity of propagation environments, persistence models provide useful context for understanding the performance of new and existing optical turbulence models.

2) CLIMATOLOGY

Rather than predicting the most recent observation of C_n^2 as in the case of persistence models, climatology models compute the mean observed C_n^2 value from the training set and use this as the prediction for all future observations. The `otbench` package implements both a standard climatology model as well as one which computes a mean value for each minute of the day. The minute climatology model uses the mean value for the given clock time seen in training as its prediction at inference time. Climatology models are included as benchmarks in the `otbench` package to provide a sense of the relative quality of new and existing models, especially in the context of specific propagation environments.

b. Data-driven models

Measurement campaigns enable researchers to fit optical turbulence models from data. These macro-meteorological models are a common approach to predicting the strength of optical turbulence from bulk measurements (Sadot and Kopeika 1992; Wang et al. 2015; Chen et al. 2019). These models are fit using data acquired from field campaigns, and are capable of generating point predictions for C_n^2 for a given measurement set. Bulk-atmospheric approaches present two key challenges; they implicitly define a set of required meteorological parameters and they present a risk of over-fitting to the local micro-climate (Wang et al. 2015; Jellen et al. 2020b). Some models, such as those in in Sadot and Kopeika (1992) or Wang et al. (2015), are parametric, while those in Jellen et al. (2020b, 2023) are non-parametric. The Gradient Boosting Regression Tree (GBRT) architecture presented in Ke et al. (2017) and the Hybrid air-water temperature difference model presented in Jellen et al. (2023) provide baseline metrics for data-driven, tree-based models as applied to the task of optical turbulence strength prediction and forecasting.

1) THE MACRO-METEOROLOGICAL MODEL

The authors of Sadot and Kopeika (1992) fit a model for C_n^2 in an over-land propagation environment at a height of 15 m:

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

where

$$f(T) = (2.0 \times 10^{-15})T \quad (1)$$

$$f(U) = (-2.5 \times 10^{-15})U + (1.2 \times 10^{-15})U^2 - (8.5 \times 10^{-17})U^3$$

$$f(RH) = (-2.8 \times 10^{-15})RH + (2.9 \times 10^{-17})RH^2 - (1.1 \times 10^{-19})RH^3$$

In equation (1), W denotes the temporal hour weight Sadot and Kopeika (1992), T denotes the temperature in K, RH denotes the relative humidity in %, and U denotes the wind speed in m s^{-1} . In order to generate predictions from this model, the dynamic range of observation presented in Sadot and Kopeika (1992) was enforced. Any measurement in which any meteorological parameters were outside of the dynamic range was dropped from the training and validation sets.

2) THE OFFSHORE MACRO-METEOROLOGICAL MODEL

The bulk-atmospheric model in Sadot and Kopeika (1992) was evaluated by the authors of Wang et al. (2015) in a coastal environment. The authors additionally fit a model for C_n^2 using data from their measurement campaign, reproduced as:

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

where

$$f(T) = 2.74 \times 10^{-16}T \quad (2)$$

$$f(U) = 3.37 \times 10^{-16}U + 1.92 \times 10^{-16}U^2 - 2.8 \times 10^{-17}U^3$$

$$f(RH) = 8.3 \times 10^{-17}RH - 2.22 \times 10^{-18}RH^2 + 1.42 \times 10^{-20}RH^3$$

In equation (2), W denotes the temporal hour weight Sadot and Kopeika (1992) T denotes the temperature in K, RH denotes the relative humidity in %, and U denotes the wind speed in m s^{-1} . As for the bulk-atmospheric model in equation (1), the dynamic range of the measurement campaign in Wang et al. (2015) was enforced when generating model predictions.

3) THE AIR-WATER TEMPERATURE DIFFERENCE MODEL

The impact of the air-water temperature difference on the strength of local optical turbulence is a feature of propagation paths over water (Frederickson et al. 2000; Jellen et al. 2020b; Chen et al. 2019). The authors of Chen et al. (2019) reproduced a model for predicting C_n^2 at ground level as a function of only the measured air-water temperature difference in °C:

$$C_n^2 = (2.05\Delta T^2 + 2.37\Delta T + 1.58) \times 10^{-16} \quad (3)$$

In equation (3), all observations in which both the air temperature and water temperature were available result in a prediction for C_n^2 (Chen et al. 2019). These predictions were scaled to the task-appropriate height using the approach described in (Jellen et al. 2023).

4) THE GBRT MODEL

The traditional Gradient Boosting Decision Tree (GBDT) architecture leverages an ensemble of decision trees to classify a sample using its features. The GBRT is a natural extension of this architecture to the regression context. As applied to optical turbulence the model seeks to predict C_n^2 from available meteorological and oceanographic data. Each constituent tree within the ensemble learns to predict a portion of the residual error iteratively improving the overall model's accuracy (Ke et al. 2017).

The GBRT's strength lies in its ability to capture complex non-linear relationships between meteorological variables and C_n^2 , as well as its robustness to missing measurements. The specific implementation presented in Ke et al. (2017) uses the mean squared error between predicted and actual values in computing gradients.

5) THE HYBRID AIR-WATER TEMPERATURE DIFFERENCE (AWT) MODEL

The hybrid model combines the air-water temperature difference model with a GBRT component. The GBRT focuses on learning the relationship between local meteorological data and the residuals of the baseline air-water temperature difference model in equation (3). By combining the predictions of both components, the hybrid model aims to provide more accurate C_n^2 estimations than the baseline model alone, especially in situations where local measurements deviate from the

conditions under which the baseline model was developed. The GBRT component serves to adapt the model to local conditions, with the objective of improving performance relative to the baseline model or a pure-GBRT model (Jellen et al. 2023).

c. Deep learning models

Deep learning models are an area of active research in the field of optical turbulence modeling (Vorontsov et al. 2020; Wang and Basu 2016). These models may offer improved performance over prior approaches, especially in contexts where high-frequency or long-term data is available for training (Vorontsov et al. 2020). The potential applicability of deep learning models for optical turbulence strength modeling motivates their inclusion as baselines in `otbench`. The `otbench` package includes a set of utilities to enable deep learning model development in forecasting and regression tasks. These utilities are built using the PyTorch framework (Paszke et al. 2019). Through the inclusion of these tools, `otbench` serves as a starting point for researchers interested in developing novel deep learning models.

1) THE RECURRENT NEURAL NETWORK MODEL

A basic, single-module recurrent neural network (RNN) architecture based on the implementation described in Elman (1990), is available for both regression and forecasting across all tasks. Optical turbulence data sets include sequential measurements of C_n^2 and other relevant meteorological parameters. This architecture has been applied for other time-series modeling tasks; while deeper models with effective hyper-parameter tuning may outperform the basic RNN model in `otbench`, this model presents a minimum baseline for comparison against future approaches.

6. Baseline model performance

a. Regression

Existing bulk-atmospheric models often focus on the challenge of predicting C_n^2 from meteorological measurements (Sadot and Kopeika 1992; Wang et al. 2015; Wang and Basu 2016; Chen et al. 2019). These models are fit using data acquired from field campaigns and are capable of generating point predictions for C_n^2 for a given measurement set. Under the `otbench` package, these are implemented as regression, and require the same set of meteorological parameters used

TABLE 6. Benchmark model performance for the `mlo_cn2` regression task, removing missing measurements.

Model	RMSE	MAE	MAPE	R^2	Valid predictions (total possible)
Macro Meteorological	0.714	0.587	0.044	0.021	813 (2449)
Offshore Macro Meteorological	0.816	0.622	0.044	0.040	2274 (2449)
Minute Climatology	0.504	0.384	0.028	0.538	2449 (2449)
Climatology	0.661	0.531	0.038	0.0	2449 (2449)
GBRT	0.212	0.154	0.011	0.910	2449 (2449)
RNN	0.336	0.213	0.015	0.761	2449 (2449)

when the model was initially fit. Some of these models leverage solar features, such as the temporal hour or temporal hour weight, derived from the local time of observation and the local sunrise time (Sadot and Kopeika 1992; Wang et al. 2015; Jellen et al. 2020b, 2023).

1) `MLO_CN2`

The two regression tasks implemented using the `mlo_cn2` data set include benchmarks for climatology, bulk-atmospheric, and data driven models. As the atmospheric boundary layer was not over water, the models in Chen et al. (2019) and Jellen et al. (2023) are not included as benchmarks for these tasks. The performance of each benchmark model, reported from $\log_{10} C_n^2$, is described in Table 6 for the task in which missing measurements are removed, and Table 7 for the task in which they are not.

The evaluation results in Table 6 reinforce the need for further development of ML and deep learning models for optical turbulence modeling. The Minute Climatology model outperforms the two bulk-atmospheric models, indicating that they may be over-fit to the propagation environments in which they were developed, or that they may be inapplicable for the conditions at Mona Loa during the study. This finding highlights the value of `otbench` as a tool for model evaluation and comparison; differential performance for a given modeling approach on a regression task requires outperforming the climatology model. The RNN model performed adequately in the regression task, demonstrating improvement over all statistical baselines. Additional architectural improvements and wider input windows may further improve the model’s performance, however, the results in Table 6 provide a minimum baseline for deep learning models as applied to this MLO

TABLE 7. Benchmark model performance for the mlo_cn2 regression task, full.

Model	RMSE	MAE	MAPE	R^2	Valid predictions (total possible)
Macro Meteorological	0.620	0.514	0.039	0.032	862 (2816)
Offshore Macro Meteorological	0.830	0.632	0.044	0.024	2641 (2816)
Minute Climatology	0.586	0.441	0.032	0.427	2816 (2816)
Climatology	0.665	0.530	0.038	0.0	2816 (2816)
GBRT	4.503	3.147	0.230	0.065	2816 (2816)
RNN	0.548	0.413	0.030	0.343	2816 (2816)

C_n^2 regression task. The random forest model used as a baseline in Jellen et al. (2023) outperforms both the RNN and climatology models, demonstrating substantial prediction accuracy with an R^2 of 0.897, indicating a high level of explained variance in measured $\log_{10} C_n^2$ at 15 m. These results, with the exception of the performance of the un-tuned, baseline GBRT model, hold for the task in which missing measurements are not removed from the data set in Table 7.

In Table 7, the RNN model again outperformed the baseline climatology model, despite the presence of missing measurements in the data. The GBRT model performs poorly in the case in which missing measurements are present across observations in the training and test sets. The dramatic drop in model performance could result from the lack of adjustment to address patterns in missingness observed in the training data.

2) USNA_CN2_SM

The scintillometer link at the United States Naval Academy is located across the Severn River, with well over 95% of the total link length over water (Jellen et al. 2021). The water temperature is available as a feature, enabling two additional models over those in 1. The air-water temperature difference model is referenced in Chen et al. (2019), while the hybrid model is introduced in (Jellen et al. 2023). All benchmark regression models are evaluated in Table 8.

Under the usna_cn2_sm regression task, the Offshore Macro-meteorological model presented in Wang et al. (2015), the random forest baseline from Jellen et al. (2020a), the RNN model, and the Hybrid AWT model introduced in Jellen et al. (2023) all outperformed the climatology baselines.

TABLE 8. Benchmark model performance for the usna_cn2_sm regression task.

Model	RMSE	MAE	MAPE	R^2	Valid predictions (total possible)
Air-water Temperature Difference	0.910	0.788	0.056	0.267	4080 (4080)
Macro Meteorological	0.864	0.685	0.047	0.268	3653 (4080)
Offshore Macro Meteorological	0.178	0.126	0.009	0.058	421 (4080)
Minute Climatology	0.452	0.362	0.026	0.126	4080 (4080)
Climatology	0.480	0.382	0.027	0.0	4080 (4080)
Hybrid AWT	0.303	0.218	0.015	0.583	4080 (4080)
GBRT	0.299	0.220	0.015	0.584	4080 (4080)
RNN	0.375	0.281	0.020	0.451	4080 (4080)

TABLE 9. Benchmark model performance for the usna_cn2_lg regression task.

Model	RMSE	MAE	MAPE	R^2	Valid predictions (total possible)
Air-water Temperature Difference	1.046	0.892	0.064	0.109	458687 (488306)
Macro Meteorological	1.217	1.045	0.072	0.015	312367 (488306)
Offshore Macro Meteorological	0.675	0.533	0.037	0.0	317350 (488306)
Minute Climatology	0.625	0.502	0.036	0.021	488306 (488306)
Climatology	0.632	0.508	0.036	0.0	488306 (488306)
Hybrid AWT	0.458	0.347	0.024	0.480	458687 (488306)
GBRT	1.340	0.683	0.048	0.029	488306 (488306)
RNN	0.530	0.400	0.028	0.362	488306 (488306)

When compared to the results in Table 7, the Minute Climatology model in Table 8 showed much lower performance improvement over the Climatology model.

3) USNA_CN2_LG

The usna_cn2_lg data set enables long-term performance evaluation for regression models. All models applicable to the regression task for the usna_cn2_sm data set also apply for the regression task built from the large data set.

As with the `usna_cn2_sm` data set, the minute-level and general climatology models demonstrated similar performance. This may be due to the weaker diurnal cycle described in (Jellen et al. 2020b). Without hyper-parameter tuning, the random forest model used as a baseline in Jellen et al. (2023) does not outperform the climatology baselines on the validation set. This could be indicative of over-fitting when applied over longer time frames, especially in the context of the model's performance for the regression task in the `usna_cn2_sm` case. The RNN model's performance is reasonable, but lower than the hybrid air-water temperature difference presented in (Jellen et al. 2023).

b. Forecasting

The challenge of forecasting optical turbulence strength differs from regression in that it allows benchmark and future models access to prior measurements of C_n^2 when predicting C_n^2 at the forecast horizon. As a result, not all regression models are applicable for forecasting tasks. The package currently implements the climatology baseline, as well as a linear forecasting model based on least-squares regression, and a model which predicts the mean C_n^2 observed within the inference window. These statistical baselines, coupled with the data-driven and deep learning models described in Section 5.

1) `MLO_CN2`

The MLO C_n^2 forecasting task allows benchmark models to observe the most recent 12 observations (1 hour), including measured C_n^2 , with the objective of predicting C_n^2 6 observations (30 minutes) in the future. The performance of each baseline forecasting model is described in Table 10.

In the forecasting context, these benchmark models establish a baseline of performance which future models may improve upon. The persistence and climatology models demonstrate lower forecasting skill than the model which predicts the mean C_n^2 observed in the inference window. The rapid fluctuation in C_n^2 often observed in the boundary layer induces error across all baseline models (Sadot and Kopeika 1992; Wang et al. 2015). Despite the long forecast horizon of 30 minute and the 15 m observation height, these baseline models generate acceptable direct forecasts for C_n^2 . This performance may, in part, be attributable to the relatively strong diurnal pattern in the MLO

TABLE 10. Benchmark model performance for the mlo_cn2 forecasting task.

Model	RMSE	MAE	MAPE	R^2	Valid predictions (total possible)
Linear	0.930	0.686	0.049	0.071	2432 (2432)
Mean Window	0.481	0.353	0.025	0.509	2432 (2432)
Persistence	1.227	1.089	0.077	0.0	2432 (2432)
Minute Climatology	0.551	0.417	0.030	0.447	2432 (2432)
Climatology	0.658	0.528	0.038	0.0	2432 (2432)
GBRT	0.428	0.318	0.023	0.652	2432 (2432)
RNN	0.581	0.439	0.032	0.403	2432 (2432)

TABLE 11. Benchmark model performance for the usna_cn2_sm forecasting task.

Model	RMSE	MAE	MAPE	R^2	Valid predictions (total possible)
Linear	0.358	0.205	0.014	0.529	4074 (4074)
Mean Window	0.182	0.102	0.007	0.847	4074 (4074)
Persistence	0.821	0.694	0.048	0.0	4074 (4074)
Minute Climatology	0.453	0.362	0.026	0.120	4074 (4074)
Climatology	0.480	0.382	0.027	0.0	4074 (4074)
GBRT	0.160	0.082	0.006	0.881	4074 (4074)
RNN	0.187	0.120	0.009	0.861	4074 (4074)

C_n^2 study. Both the RNN and GBRT models out-performed the statistical baselines, demonstrating reasonable prediction accuracy in direct forecasting. The relative performance improvement was lower for the forecasting task than for the corresponding regression task in Table 6. Auto-regressive approaches may further improve these baselines.

2) USNA_CN2_SM

The usna_cn2_sm data set was used to develop a short-term forecasting task in which the next observation of C_n^2 (6 minutes in the future) is predicted using the last 6 observations (36 minutes) during model inference. The diurnal pattern over the Severn River was not as strong as that observed during the MLO C_n^2 study, and fluctuations C_n^2 were more rapid due to the lower height of observation and other attributes of the propagation environment. The same baseline models used in Table 10 are evaluated in Table 11.

As for the MLO C_n^2 forecasting task, the model which predicts the mean value of C_n^2 observed in the inference window outperforms other persistence and climatology baselines. Due to the short-term nature of forecasts in this task, baseline model error in Table 11 is lower across all evaluation metrics when compared to the results in Table 10. The GBRT model outperforms the mean window model slightly, while performance for the RNN model is comparable to the strongest baseline. The low performance improvement of data-driven and deep learning approaches relative to simpler statistical baselines motivates investigation into architectural improvements for the forecasting context.

7. Discussion

The `otbench` package enables robust evaluation and comparison of models for predicting and forecasting optical turbulence strength. The implementation of baseline models allows researchers to compare the performance of new models against statistical, bulk-atmospheric, data-driven, and deep learning models developed across prior studies (Rasp et al. 2020; Sadot and Kopeika 1992; Wang et al. 2015; Jellen et al. 2023). By demonstrating the applicability of both traditional and ML-based models, `otbench` establishes a bridge between established methodologies and emerging ML techniques in the field of optical turbulence research. Differential performance above these benchmarks can also be assessed across more than one field campaign. Researchers can compare their approach across three major measurement campaigns; the package is easily extensible for new data sets and tasks.

As seen in Table 6, Table 8, and Table 9, data-driven and deep learning models can perform well in predicting the intensity of optical turbulence measured as C_n^2 from meteorological data. These results further indicate that baseline models such as the minute-level climatology model perform well, better than the traditional climatology model, in environments with stronger diurnal patterns. This could be further investigated by incorporating a data set and associated regression task captured using a boundary-layer link over a desert environment. In the forecasting context, the model which predicts the mean of C_n^2 measurements within the inference window acts as a strong benchmark for the development of future models. This model outperforms both the minute-level climatology model and the general climatology model for both forecasting tasks, as seen in Table 10

and Table 11. Data-driven and deep learning models show promising results in predicting C_n^2 from meteorological data, highlighting the potential of ML to advance optical turbulence modeling.

While the current focus of the `otbench` framework is in establishing standardized evaluation, its future development will emphasize the integration and evaluation of more sophisticated ML techniques. This evolution will facilitate the exploration of new architectures, algorithms, and feature engineering approaches for optical turbulence modeling, potentially unlocking new insights into the complex dynamics of atmospheric turbulence.

8. Conclusions

The accurate prediction and forecasting of optical turbulence is of paramount importance for the deployment of reliable laser-based systems in real-world environments. The `otbench` package addresses the long-standing need for a standardized evaluation framework and benchmark data sets in this field. By providing a comprehensive platform for model development, evaluation, and comparison, `otbench` empowers researchers to build upon existing knowledge and explore the full potential of ML techniques for optical turbulence modeling.

The data sets which `otbench` makes available allow for the investigation and development of optical turbulence models across a range of operational conditions. Standardized evaluation of existing and proposed models, including both traditional and ML methods, can be objectively assessed across different data sets and tasks. These evaluations reflect the relative effectiveness of models as applied to the challenge of optical turbulence modeling, and the critical need for further improvement. The data, source code, and package are available at <https://github.com/cdjellen/otbench>.

Acknowledgments. The authors would also like to thank the Waterfront Readiness Center and the United States Naval Academy Mechanical Engineering, Electrical Engineering, Oceanography, and Mathematics Departments for their assistance in conducting the measurement campaign. We are grateful to the NCAR and UCAR data archive teams for making the MLO C_n^2 data set available, and to Gordon Maclean and Steve Oncley for their development of the Mauna Loa Seeing Study. The authors would also like to thank the United States Naval Academy Trident Scholar for their support of this research effort.

Data availability statement. The source code and data sets for the otbench package are available at <https://github.com/cdjellen/otbench>. Data from the MLO C_n^2 study from NCAR/EOL is available at <https://data.eol.ucar.edu/dataset/160.007>.

References

- Andrews, L. C., and R. L. Phillips, 2005: *Laser beam propagation through random media*.
- Barnett, S., J. Blau, P. Frederickson, and K. Cohn, 2022: Measurements and modeling of optical turbulence in the coastal environment. *Applied Sciences*, **12** (10), <https://doi.org/10.3390/app12104892>.
- Center for Operational Oceanographic Products and Services, 2023: Annapolis, md - station id: 8575512. National Oceanic and Atmospheric Administration, URL <https://tidesandcurrents.noaa.gov/stationhome.html?id=8575512>.
- Chen, L.-H., Z. Liu, and D. Chen, 2019: Climatological analysis of the seeing at fuxian solar observatory. *Research in Astronomy and Astrophysics*, **19** (1), 015.
- Elman, J. L., 1990: Finding structure in time. *Cognitive science*, **14** (2), 179–211.
- Frederickson, P. A., K. L. Davidson, C. R. Zeisse, and C. S. Bendall, 2000: Estimating the refractive index structure parameter (cn2) over the ocean using bulk methods. *Journal of applied meteorology*, **39** (10), 1770–1783.
- Hu, W., M. Fey, M. Zitnik, Y. Dong, H. Ren, B. Liu, M. Catasta, and J. Leskovec, 2020: Open graph benchmark: Datasets for machine learning on graphs. *Advances in neural information processing systems*, **33**, 22 118–22 133.

- Jellen, C., J. Burkhardt, C. Brownell, and C. Nelson, 2020a: Machine learning informed predictor importance measures of environmental parameters in maritime optical turbulence. *Applied Optics*, **59** (21), 6379–6389.
- Jellen, C., C. Nelson, C. Brownell, and J. Burkhardt, 2024: Long-term measurement and characterization of boundary layer optical turbulence. *J. Opt. Soc. Am. A*, **41** (6), B65–B72, <https://doi.org/10.1364/JOSAA.520980>.
- Jellen, C., C. Nelson, C. Brownell, J. Burkhardt, and M. Oakley, 2020b: Measurement and analysis of atmospheric optical turbulence in a near-maritime environment. *IOP SciNotes*, **1** (2), 024 006.
- Jellen, C., C. Nelson, J. Burkhardt, and C. Brownell, 2023: Hybrid optical turbulence models using machine-learning and local measurements. *Applied Optics*, **62** (18), 4880–4890.
- Jellen, C., M. Oakley, C. Nelson, J. Burkhardt, and C. Brownell, 2021: Machine-learning informed macro-meteorological models for the near-maritime environment. *Appl. Opt.*, **60** (11), 2938–2951, <https://doi.org/10.1364/AO.416680>.
- Ke, G., Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, 2017: Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, **30**.
- Maclean, G., and S. Oncley, 2006: Mauna loa seeing study. NCAR In-Situ Sensing Facility, URL https://archive.eol.ucar.edu/docs/isf/projects/MLO_CN2/.
- Mahon, R., C. I. Moore, M. S. Ferraro, W. S. Rabinovich, and P. A. Frederickson, 2020: Comparison of maritime measurements of cn2 with navslam model predictions. *Applied Optics*, **59** (33), 10 599–10 612.
- National Data Buoy Center, 2023: Station tplm2 - thomas point, md. National Data Buoy Center, URL https://www.ndbc.noaa.gov/station_page.php?station=tplm2.
- NCAR/EOL ISFS Team , 2023: Mlo_cn2: Ncar/eol isfs 5-minute surface meteorology and flux products. version 1.0. UCAR/NCAR - Earth Observing Laboratory, URL <https://data.eol.ucar.edu/dataset/160.007>, <https://doi.org/10.26023/CQR2-TQJ9-AH10>.

- Paszke, A., and Coauthors, 2019: Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, **32**.
- Pierzyna, M., R. Saathof, and S. Basu, 2023: Pi-ml: A dimensional analysis-based machine learning parameterization of optical turbulence in the atmospheric surface layer. *arXiv preprint arXiv:2304.12177*.
- Rasp, S., P. D. Dueben, S. Scher, J. A. Weyn, S. Mouatadid, and N. Thuerey, 2020: Weatherbench: a benchmark data set for data-driven weather forecasting. *Journal of Advances in Modeling Earth Systems*, **12** (11), e2020MS002 203.
- Sadot, D., and N. S. Kopeika, 1992: Forecasting optical turbulence strength on the basis of macroscale meteorology and aerosols: models and validation. *Optical Engineering*, **31** (2), 200–212.
- Vorontsov, A. M., M. A. Vorontsov, G. A. Filimonov, and E. Polnau, 2020: Atmospheric turbulence study with deep machine learning of intensity scintillation patterns. *Applied Sciences*, **10** (22), 8136.
- Wang, H., B. Li, X. Wu, C. Liu, Z. Hu, and P. Xu, 2015: Prediction model of atmospheric refractive index structure parameter in coastal area. *Journal of Modern Optics*, **62** (16), 1336–1346.
- Wang, Y., and S. Basu, 2016: Using an artificial neural network approach to estimate surface-layer optical turbulence at mauna loa, hawaii. *Optics letters*, **41** (10), 2334–2337.