

PAPER • OPEN ACCESS

Data-driven prediction of mean wind turbulence from topographic data

To cite this article: B Morais da Costa *et al* 2021 *IOP Conf. Ser.: Mater. Sci. Eng.* **1201** 012005

View the [article online](#) for updates and enhancements.

You may also like

- [Development of a method for evaluation of measures used for winterization of offshore facilities and units](#)
S R Jacobsen, S Viddal, K G Dørum *et al.*
- [Testing of a new transport and installation method for offshore wind turbines](#)
J Haugvaldstad and O T Gudmestad
- [Investigation of updating methods for probability-informed inspection planning for offshore structures](#)
G Ersdal and N Oma



The Electrochemical Society
Advancing solid state & electrochemical science & technology

241st ECS Meeting

May 29 – June 2, 2022 Vancouver • BC • Canada

Extended abstract submission deadline: Dec 17, 2021

Connect. Engage. Champion. Empower. Accelerate.
Move science forward



Submit your abstract



Data-driven prediction of mean wind turbulence from topographic data

B Morais da Costa^{1,2*}, J Þ Snæbjörnsson^{2,3}, O A Øiseth⁴, J Wang^{1,2} and J B Jakobsen²

¹ Norwegian Public Roads Administration, Bergelandsgata 30, 4012 Stavanger.

² Department of Mechanical and Structural Engineering and Materials Science, University of Stavanger, Kristine Bonnevis vei 22, 4021 Stavanger, Norway.

³ Department of Engineering, Reykjavik University, Menntavegur 1, 102 Reykjavík, Iceland.

⁴ Department of Structural Engineering, Norwegian University of Science and Technology, Richard Birkelands vei 1A, 7491 Trondheim, Norway.

* Contact author: bercos@vegvesen.no

Abstract. This study presents a data-driven model to predict mean turbulence intensities at desired generic locations, for all wind directions. The model, a multilayer perceptron, requires only information about the local topography and a historical dataset of wind measurements and topography at other locations. Five years of data from six different wind measurement mast locations were used. A k-fold cross-validation evaluated the model at each location, where four locations were used for the training data, another location was used for validation, and the remaining one to test the model. The model outperformed the approach given in the European standard, for both performance metrics used. The results of different hyperparameter optimizations are presented, allowing for uncertainty estimates of the model performances.

1. Introduction

Wind turbulence, in the atmospheric boundary layer, is an important phenomenon in the design of civil structures for both static and dynamic wind loads, and for the safe operation of transport vehicles. It arises from both mechanical and thermal sources. Frictional forces between the moving air and the Earth's surface are the main drivers of atmospheric turbulence and are closely linked to the local topography. Thermal sources such as surface heating/cooling and downbursts can also cause turbulence in the atmosphere by convection.

Measuring the wind properties at some desired locations can be challenging, despite promising advances in remote sensing [1-4]. Cheynet et al. [5] showed a high heterogeneity of wind turbulence in a fjord with the wind direction, which can significantly impact the design of wind-sensitive bridges and other man-made structures. In these situations, wind measurements, when available, are often only found at nearby locations. If there is enough diversity in the topography of the available measurement locations and sufficient wind data is available, it is in principle possible to use machine learning to learn the complex effects that the upstream topography has on the wind turbulence.

Artificial neural networks (ANN) (see e.g. [6, 7]), can be of different types. Among them, multilayer perceptrons [8, 9] have been used in many problems in atmospheric sciences [10]. They have been used



to e.g. predict wind speeds from ocean surface images [11], and effectively identify topographic features such as water bodies, hills and vegetation [12]. They are thus deemed adequate for the simplified vector inputs used in this study, despite a broader support for e.g. convolutional neural networks and transformers in more challenging computer vision tasks [13, 14]. To the authors knowledge, this study is the first attempt to use topographic information to predict mean wind turbulence intensities at new locations, without explicitly parametrizing the topography. Parametric models representing terrain effects are inherently imperfect and are based on numerous simplifications and difficult assessments in an attempt to systematically represent a complex terrain. They were previously proposed in e.g. the Eurocode [15], Engineering Science Data Unit [16] and Bitsuamlak et al. [17]. Other studies [18-20] model the dependencies between wind measurements at different locations and predict wind speeds, but are unable to predict mean wind characteristics at new locations where no measurements were available, given only information about the local topography. Bodini et al. [21] predict the turbulent kinetic energy dissipation rate while condensing the effects of the upstream topography into two variables, namely the standard deviation of the terrain elevation and the mean vegetation height, but also test their model at previously trained locations.

The model developed in this study is trained, validated and tested using measured along-wind turbulence intensities that are averaged within 1-degree-wide wind direction sectors, here denoted sectoral averages, and the topographic data associated with each sector, for each measurement mast location. The model hyperparameters were optimized after each iteration of a so-called k-fold cross-validation, and uncertainty estimates were provided for the model performance on each tested location.

2. Data description

2.1. Wind measurement data

Five years of wind data, between 2015 and 2020, from six measurement masts in the region around the Bjørnafjord, in Norway, are used. The locations and names of these masts are shown in Figure 1.

Each mast has 3 sonic anemometers (model: *Gill WindMaster Pro*) that measured the three components of the wind with a sampling frequency of 10 Hz. The anemometers are located at 13, 33 and 48 meters above ground. To avoid measurements affected by smaller nearby obstacles such as trees and buildings, which are not represented in the topographic data, only the data recorded at 48 m height was used, for simplicity. Thus, the turbulence intensities being predicted at the different locations also refer to a 48 m height above ground. The data is pre-processed to address faulty and missing data. An outlier detection is performed through a Z-score analysis, where the 99.99% most probable data is kept.

For each 10-minute interval in the five-year period, the mean wind speed U , the mean wind direction and the along-wind turbulence intensity I_u are recorded for each anemometer at 48 m height, when available. A threshold of 5 m/s was adopted and observations with smaller mean wind speeds were discarded. High threshold values require more data but help to remove turbulence observations that are not likely governed by friction, but by e.g. local thermal effects.

2.2. Topographic data

The Norwegian mapping authority provides freely accessible Digital Terrain Models of Norway [22]. A 10×10 meter resolution model was used (DTM 10), consistently represented in the map projection system UTM 33. For each mast and for each 1-degree-wide wind sector, a 10 km long upstream terrain profile aligned with the wind was obtained. Note that a 10 km fetch is also suggested in NS-EN 1991-1-4:2005+NA:2009 NA.4.3.2(2) (901.1). The heights above sea level of 45 points along the profile at the upstream distances $X = [0, 10, 30, 60, 100, 150, \dots, 9900]$ (meters) were collected into a normalized terrain profile vector Z , where for each single point a min-max normalization is performed from that point's extreme values (for all masts and directions), as exemplified in Figure 2. Note the linearly increasing distance between points. This decrease in resolution assumes that, far upstream, only larger topographic features still affect turbulence (see e.g. [15], NA.4.3.2(2) (901.2)). Different sizes of X , between 15 and 60, were also tested, with roughly similar results. To consider the effect of the different

categories of terrain roughness, a vector R was added to the data used. Two categories were considered, sea and ground, normalized into a binary vector, but more terrain categories could be included.



Figure 1. Locations and adopted names of the wind measurement masts used.

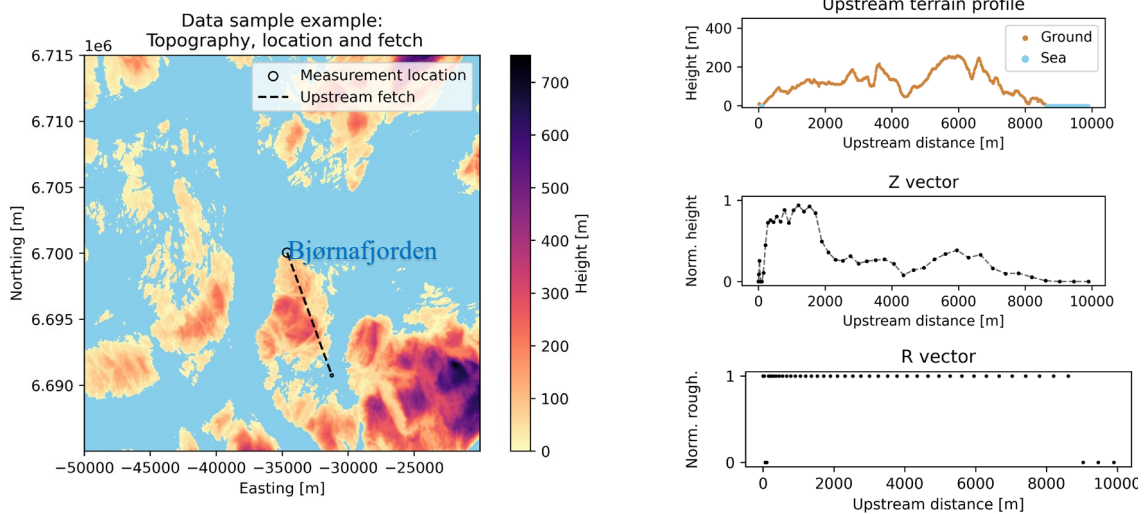


Figure 2. Example of an elevation (Z) and roughness (R) data sample collection, given a location and a mean wind direction, associated with an upstream fetch and terrain profile.

3. Methodology

3.1. Artificial neural network

An artificial neural network (ANN) was established using *PyTorch* (v.1.9.0) (a *Python* library for deep learning). A multilayer perceptron arrangement was used, whose representation is shown in Figure 3. The ANN predicts the sectoral/directional averages of the along-wind turbulence intensities \bar{I}_u , i.e., the mean value of all I_u within each 1-degree wide wind sector, at each wind mast, at 48 m above ground. A k-fold cross-validation method is used where the data is divided into six folds and where each fold corresponds to the data of one measurement mast location. This forces the model to predict turbulence intensities at locations it has never “seen” before. Each fold contains up to 360 data samples, one for each wind sector. The procedure for training, validating, optimizing and testing is further detailed in Figure 3. The domain of hyperparameters investigated is described in Table 1.

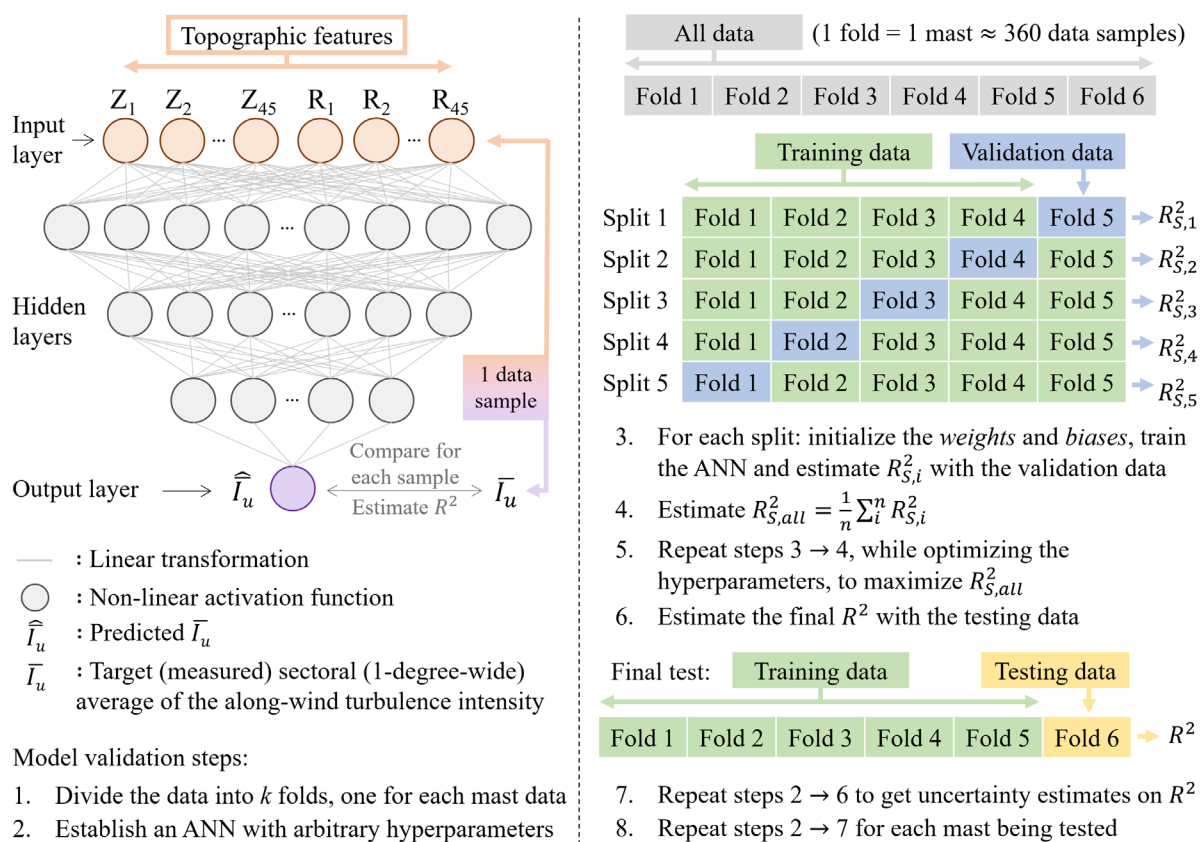


Figure 3. Representation of the multilayer perceptron and the cross-validation procedure adopted.

Table 1. Domain of hyperparameters investigated.

Hyperparameter	Sampled intervals / categories
Num. of hidden layers	0 – 4 (each having $\approx 2/3$ of the size of the preceding one)
Activation function	ReLU / Leaky ReLU / ELU
Loss function	L1 Loss / MSE Loss / Log-Cosh Loss
Num. of epochs	20 - 2000
Learning rate	10^{-3} - 1
Weight decay	10^{-5} - 1
Momentum	0 - 0.95

A min-max normalization is applied to all inputs and target outputs to improve learning and stability. The target values \bar{I}_u are compared with the predicted values \hat{I}_u through a loss function and learning is achieved by backpropagation. A batch gradient descent was found suitable due to limited data size and use of GPU-accelerated algorithms. The hyperparameters were optimized to maximize the R^2 value (coefficient of determination) of the validation data predictions, using 500 iterations with a so-called “Tree-structured Parzen Estimator Approach”. This is preferable to grid and random searches and has been shown to have a good balance between performance and computer efficiency when compared to other methods such as gaussian processes and random forests [23, 24]. Since the resultant “optimal” hyperparameters depend on the initial conditions, 20 initial sets of arbitrary hyperparameters, thus 20 different models, were used to estimate the uncertainty of the R^2 of the final testing data predictions. Lastly, when predicting the sectoral averages \bar{I}_u , instead of each 10-min occurrence of I_u , the topographic effects are better isolated and other time- and thermal-related effects can be disregarded.

3.2. Norwegian Standard - Eurocode NS-EN 1991-1-4

For comparison purposes, the along-wind turbulence intensity is also estimated following the Norwegian Standard and Eurocode NS-EN 1991-1-4 (ref. [15]). The measurement masts presented in this study are in a region with strong contrasts of terrain roughness, namely sea water (terrain cat. 0) and forests in relatively small hills (terrain cat. III). This transition in the upstream terrain roughness is considered in the Eurocode NA.4.3.2(2) (901.2.2). Different orographic effects on turbulence could also be considered. Those described in NA.4.3.3 (901.2.1) and NA.4.3.3 (901.3.2) can be applicable to some of the studied locations. However, in NA.4.4 it is not clear how to combine these effects with those from the different terrain roughnesses upstream, so only the latter ones are considered. Also, the orography factor is intended to represent isolated hills and escarpments, not undulating and mountainous regions.

To consider the upstream roughness heterogeneity, the upstream terrain is divided into two continuous patches of either terrain category 0 or III. The length of the two patches and the location of the transition between them was found iteratively for each mast and wind direction, by minimizing the number of misclassifications when compared to the original R vector.

4. Results and discussion

The main results of the data-driven analysis are presented in Figure 4, Figure 5 and Figure 6.

In Figure 4, the predictions of one ANN model, per location, are plotted. The plotted models were those that had their performance (R^2) closest to the average performance of all models for a given location (dark red dots in Figure 5). Displaying only the best performing ANN models would lead to bias, due to a regression to the mean of future dataset test performances, and is thus avoided. Contour and line plots are shown for each mast location. The contour plots show the upstream topography for each mean wind direction, with the same resolution as given in the input data for the Z and R vectors (see Section 2.2). A blue color is superposed to represent the sea water, with lower surface roughness. The line plots show the measurements and ANN predictions of \bar{I}_u . The sectoral averages of the mean wind speeds, \bar{U} , from the data described in Section 2.1, are also included for completeness.

Upstream hills close to the masts affect the results to a greater extent than hills further away. Long upstream fetches of water are characterized by low turbulence intensities. The ANN predictions are best at Ospøya 1 and Ospøya 2 as expected, due to the proximity (260 meters) and topographic similarity between them. Some predicted values at Ospøya 2, Landrøppynten and Nesøya seem slightly misaligned with the measurements. This can be due to local deflections of the wind direction around hills or/and due to discrepancies between reported and real anemometer orientations. At Svarvhelleholmen, the ANN underestimates turbulence for southern winds due to the inexistence of such high turbulence intensities in its training database. The Eurocode prediction also underestimates turbulence, but it could be argued that the alternative procedure in NA.4.3.3 (901.3.2) (“Lower lying construction site downstream of a hill or escarpment”) would lead to slightly higher turbulence intensities for this particular site and direction. At Synnøytangen, the presence of nearby buildings and tall trees presumably affects the measurements to some extent for some directions.

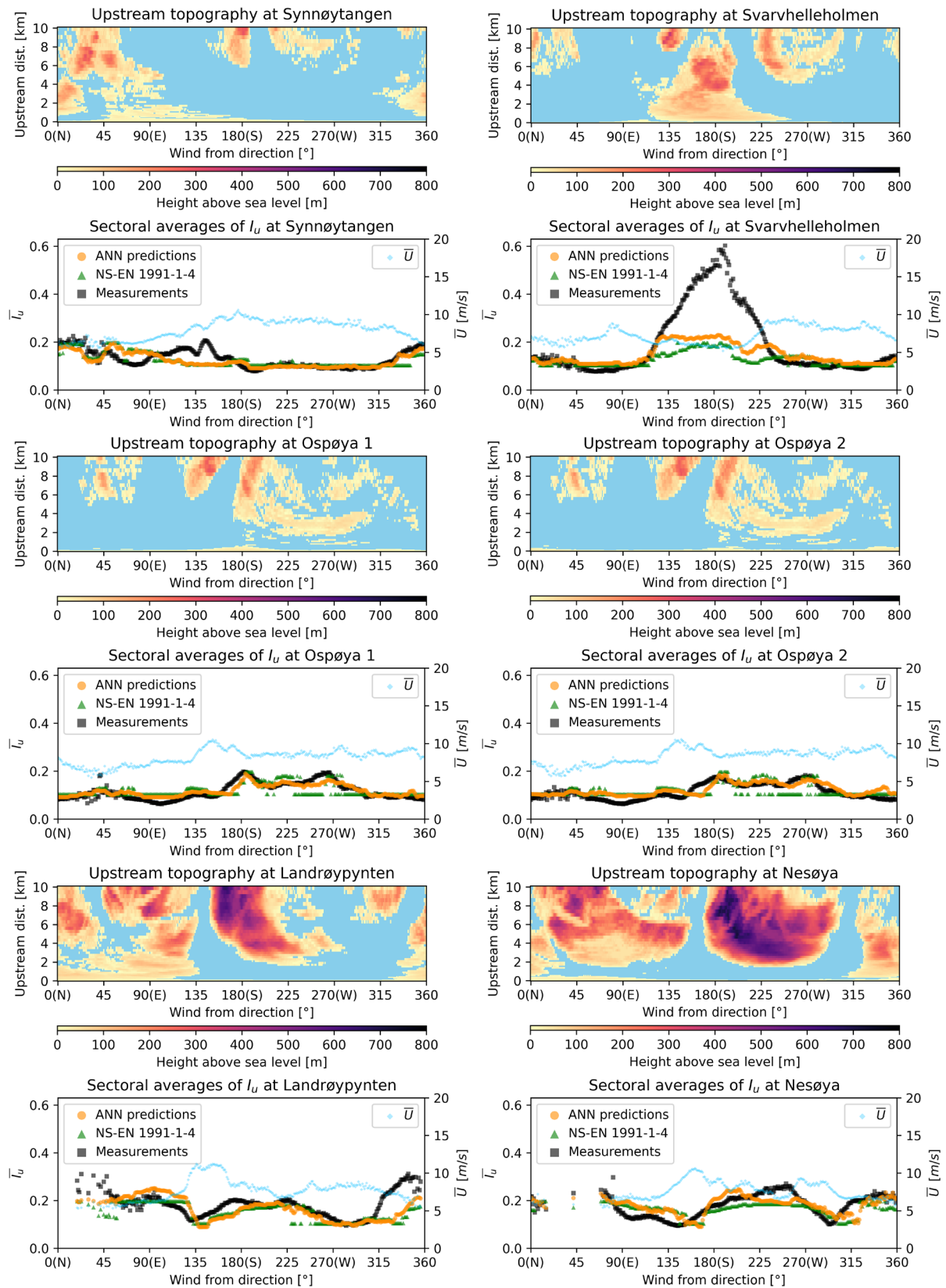


Figure 4. Upstream topography for each mast and wind direction. Sectoral (1-degree-wide) averages of I_u : measurements vs predictions of an average performing ANN vs NS-EN 1991-1-4.

In Figure 5, the R^2 values, between the predictions of all ANN models and the tested measurements, are shown as an indication of the model performances. Note that the hyperparameter optimization is a chaotic process that is dependent on the initial conditions, hence the 20 models per tested location and associated R^2 uncertainty estimates. A value of $R^2 = 1$ indicates a perfect fit, whereas $R^2 = 0$ indicates a fit that is as good as a simple average of all 360 values of \bar{I}_u (which is unknown a priori). Another performance metric, accuracy, is also included, taken as $100\% - MAPE$ (mean absolute percentage error).

In Figure 6, seven histograms show the final choices of the hyperparameters for all the different ANN models tested, after all optimization iterations were complete. It took roughly 110 hours to compute the 6 mast locations \times 20 ANN models \times 500 optimization iterations, on a laptop PC (Intel Core i7-8850H, 64 GB 2666 MHz RAM, Nvidia Quadro P4200).

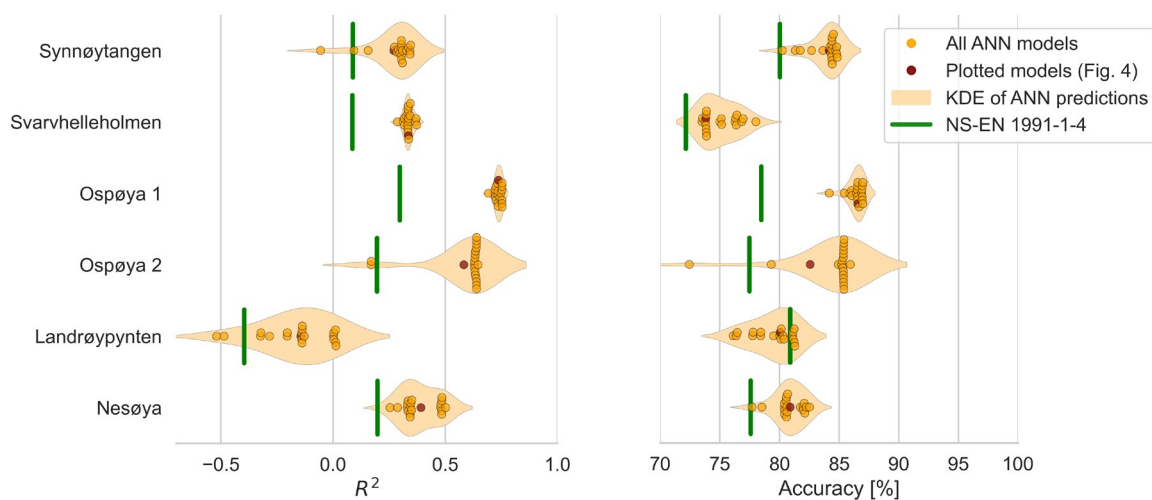


Figure 5. R^2 values and accuracies of the ANN predictions of \bar{I}_u . Kernel density estimations (KDE) are provided to help visualize the underlying uncertainties. The Eurocode (NS-EN 1991-1-4) predictions are included for comparison.

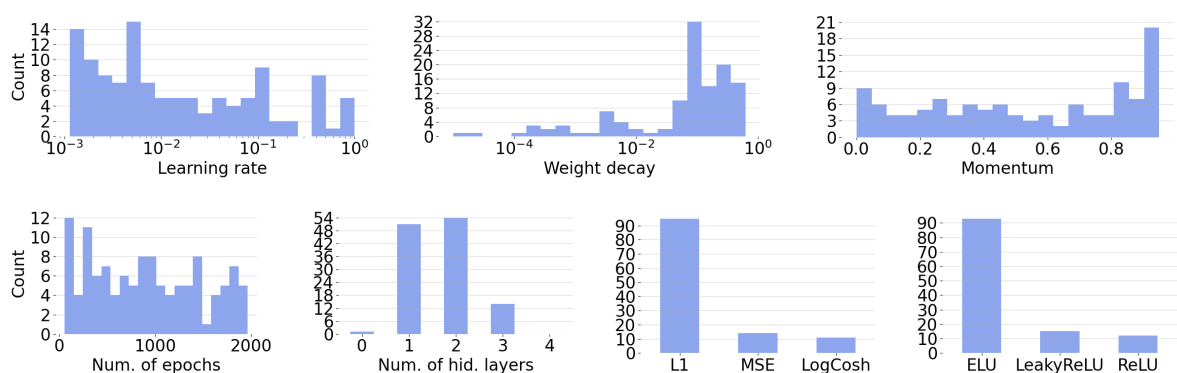


Figure 6. Histograms of the optimal hyperparameters of all ANN models.

For all masts, the ANN predictions were able to roughly capture the main trends of the mean turbulence intensities with the location and wind direction, showing overall better performances than the Eurocode predictions. Nonetheless, it remains a challenging task to accurately predict turbulence, regardless of the model adopted.

5. Conclusions

A data-driven model was developed to predict mean wind turbulence intensities for each mean wind direction in a complex terrain, where no wind measurements are available. The model consists of an artificial neural network, namely a multilayer perceptron, whose hyperparameters were systematically optimized to improve the predictions. First, a database of topographic data and measured turbulence intensities at 48 meters height above ground, at different locations, for each wind direction, was used to train the model. Each topographic data sample consisted of 45 terrain elevation points, associated with a location, a wind direction and an upstream terrain profile, plus 45 binary classifications of those points' roughnesses into "ground" or "sea". Then, the model required only the topographic data at the desired new location to predict the mean turbulence intensities at the same height above ground, for each mean wind direction.

For the six locations studied, prediction accuracies between 72% and 87% were obtained, despite the relatively small training databases with only four or five locations. The model outperformed the procedures given in the relevant standard (Eurocode NS-EN 1991-1-4), which inherently require numerous simplifications that are difficult to implement and systematize in a complex terrain. The model is simple to establish, and the suggested framework can be easily adapted to include other input features and/or to predict other wind properties.

These findings can be useful when estimating the design wind loads on structures in complex terrains as a function of the wind direction. The proof-of-concept presented could also encourage other stakeholders in establishing a comprehensive and global database, with a larger number of measurement locations and diversified topographies, which could lead to an increase in model accuracy and reliability. Such a database and model could significantly impact the design, safety and cost-effectiveness of wind sensitive structures.

6. Recommendations for further work

A few recommendations and ideas on how to expand the current work are as follows:

- Wind measurements at different heights above ground should be collected, to expand the scope of the model and capture the turbulence relationship with the height above ground.
- More terrain categories, or a continuous roughness parameter, could be directly estimated as in [25, 26], using e.g. the finer point cloud models available in [22] (0.25×0.25 m resolution).
- The crosswind and vertical turbulence intensities, often assumed to have a linear relationship with the along-wind turbulence, could be included in the model.
- Expanding the inputs to "see" a wider upstream topography, such as a $\pm 15^\circ$ sector around the wind direction, could improve the predictions and capture effects such as wind deflection around hills and the horizontal diffusion of turbulence. All-around topographies could also be considered, to capture channeling and downstream blockage effects. In the present study and limited data, this resulted in no obvious gains in accuracy.
- Convolutional neural networks and other state-of-the-art computer vision models could be used to capture the spatial information of the expanded inputs mentioned above.
- A hybrid ANN + Eurocode model could be pursued, where the Eurocode predictions could be added to the ANN inputs.
- Predefined probability density functions of wind turbulence could be predicted instead of the sectoral mean turbulence intensity. Attempts in the present study have shown that functions with more parameters resulted in a better representation of the real data, but led to worse predictions, and vice-versa, presumably due to the lack of data in some wind sectors and the small number of mast locations in the database.

Acknowledgments

This project was financially supported by the Norwegian Public Roads Administration and the Research Council of Norway. The authors would also like to acknowledge Etienne Cheynet, Hugo Moreira and Helle Fuhr for the valuable discussions.

References

- [1] Floors R, Peña A, Lea G, Vasiljević N, Simon E and Courtney M 2016 The RUNE experiment—A database of remote-sensing observations of near-shore winds *Remote Sens.* **8**, 884
- [2] Cheynet E, Jakobsen J, Snæbjörnsson J, Ágústsson H and Harstveit K 2018 Complementary use of wind lidars and land-based met-masts for wind measurements in a wide fjord. In: *Journal of Physics: Conference Series*: (IOP Publishing) p 012028
- [3] Bourassa M A, Meissner T, Cerovecki I, Chang P S, Dong X, De Chiara G, Donlon C, Dukhovskoy D S, Elya J and Fore A 2019 Remotely sensed winds and wind stresses for marine forecasting and ocean modeling *Front. Mar. Sci.* **6**, 443.
- [4] Simma M, Mjøen H and Boström T 2020 Measuring Wind Speed Using the Internal Stabilization System of a Quadrotor Drone *Drones* **4**, 23.
- [5] Cheynet E, Liu S, Ong M C, Jakobsen J B, Snæbjörnsson J and Gatin I 2020 The influence of terrain on the mean wind flow characteristics in a fjord *J. Wind Eng. Ind. Aerodyn.* **205** 104331
- [6] Rosenblatt F 1958 The perceptron: a probabilistic model for information storage and organization in the brain *Psychol. Rev.* **65**, 386.
- [7] Graupe D 2013 *Principles of artificial neural networks vol 7*: (World Scientific)
- [8] Rosenblatt F 1961 Principles of neurodynamics. perceptrons and the theory of brain mechanisms. (Cornell Aeronautical Lab Inc Buffalo NY)
- [9] Rumelhart D E, Hinton G E and Williams R J 1985 Learning internal representations by error propagation. (California Univ San Diego La Jolla Inst for Cognitive Science)
- [10] Gardner M W and Dorling S 1998 Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences *Atmos. Environ.* **32** 2627-36.
- [11] Krasnopolsky V, Breaker L and Gemmill W 1995 A neural network as a nonlinear transfer function model for retrieving surface wind speeds from the special sensor microwave imager *J. Geophys. Res.: Oceans* **100**, 11033-45.
- [12] Venkatesh Y and Raja S K 2003 On the classification of multispectral satellite images using the multilayer perceptron *Pattern Recognit.* **36**, 2161-75.
- [13] Sharif Razavian A, Azizpour H, Sullivan J and Carlsson S 2014 CNN features off-the-shelf: an astounding baseline for recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp 806-13.
- [14] Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G and Gelly S 2020 An image is worth 16x16 words: Transformers for image recognition at scale *arXiv preprint arXiv:2010.11929*.
- [15] Norwegian Standard. Eurocode 1: Actions on Structures—Part 1–4: General Actions—Wind Actions (NS-EN 1991-1-4:2005+NA:2009).
- [16] UNIT E S D 1995 Computer program for wind speeds and turbulence properties: flat or hilly sites in terrain with roughness changes. (ESDU International London)
- [17] Bitsuamlak G, Bédard C and Stathopoulos T 2007 Modeling the effect of topography on wind flow using a combined numerical–neural network approach *J. Comput. Civ. Eng.* **21** 384-92
- [18] Brahim T 2019 Using artificial intelligence to predict wind speed for energy application in Saudi Arabia *Energies* **12**, 4669.
- [19] Velo R, López P and Maseda F 2014 Wind speed estimation using multilayer perceptron *Energy Convers. Manage.* **81**,1-9.
- [20] Ehsan M A 2021 Predictive models for wind speed using artificial intelligence and copula. In: *Department of Electrical and Computer Engineering*: (University of the District of Columbia)
- [21] Bodini N, Lundquist J K and Optis M 2020 Can machine learning improve the model representation of turbulent kinetic energy dissipation rate in the boundary layer for complex terrain? *Geosci. Model Dev.* **13**, 4271-85.
- [22] Kartverket Høydedata (<https://hoydedata.no/>).
- [23] Bergstra J, Bardenet R, Bengio Y and Kégl B 2011 Algorithms for hyper-parameter optimization *Adv. Neur. In.* **24**

- [24] Akiba T, Sano S, Yanase T, Ohta T and Koyama M 2019 Optuna: A next-generation hyperparameter optimization framework. In: *Proceedings of the 25th ACM SIGKDD Int conf on knowledge discovery & data mining*, pp 2623-31
- [25] Mah J, Samson C, McKinnon S D and Thibodeau D 2013 3D laser imaging for surface roughness analysis *Int. J. Rock Mech. Min. Sci.* **58** 111-7.
- [26] Mills G and Fotopoulos G 2013 On the estimation of geological surface roughness from terrestrial laser scanner point clouds *Geosphere* **9** 1410-6.