https://doi.org/10.5194/gmd-2020-16
Preprint. Discussion started: 21 April 2020

Geoscientific Model Development Discussions

Can machine learning improve the model representation of TKE dissipation rate in the boundary layer for complex terrain?

Nicola Bodini[1,2], Julie K. Lundquist[1,2], and Mike Optis[2]

[1]Department of Atmospheric and Oceanic Sciences, University of Colorado Boulder, Boulder, Colorado, USA
[2]National Renewable Energy Laboratory, Golden, Colorado, USA

Correspondence: Nicola Bodini (nicola.bodini@nrel.gov)

Abstract. Current turbulence parameterizations in numerical weather prediction models at the mesoscale assume a local equilibrium between production and dissipation of turbulence. As this assumption does not hold at fine horizontal resolutions, improved ways to represent turbulent kinetic energy (TKE) dissipation rate (ϵ) are needed. Here, we use a 6-week data set of turbulence measurements from 184 sonic anemometers in complex terrain at the Perdigão field campaign to suggest improved representations of dissipation rate. First, we demonstrate that a widely used Mellor, Yamada, Nakanishi, and Niino (MYNN) parameterization of TKE dissipation rate leads to a large inaccuracy and bias in the representation of ϵ. Next, we assess the potential of machine-learning techniques to predict TKE dissipation rate from a set of atmospheric and terrain-related features. We train and test several machine-learning algorithms using the data at Perdigão, and we find that multivariate polynomial regressions and random forests can eliminate the bias MYNN currently shows in representing ϵ, while also reducing the average error by up to 30%. Of all the variables included in the algorithms, TKE is the variable responsible for most of the variability of ϵ, and a strong positive correlation exists between the two. These results suggest further consideration of machine-learning techniques to enhance parameterizations of turbulence in numerical weather prediction models.

1 Introduction

While turbulence is an essential quantity that regulates many phenomena in the atmospheric boundary layer (Garratt, 1994), numerical weather prediction models are not capable of fully resolving it. Instead, they rely on parameterizations to represent some of the turbulent processes. Investigations into model sensitivity have shown that out of the various parameterizations

https://doi.org/10.5194/gmd-2020-16
Preprint. Discussion started: 21 April 2020


Geoscientific Model Development Discussions

currently used in mesoscale models, that of turbulent kinetic energy (TKE) dissipation rate (ϵ) has the greatest impact on the accuracy of model predictions of wind speed at wind turbine hub height (Yang et al., 2017; Berg et al., 2018).

Current boundary layer parameterizations of ϵ in mesoscale models assume a local equilibrium between production and dissipation of TKE. While this assumption is generally valid for homogeneous and stationary flow (Albertson et al., 1997), as the horizontal grid resolution of mesoscale models is constantly pushed toward finer scales thanks to the increase of the computing resource capabilities, the theoretical bases of this assumption are violated. In fact, turbulence produced within a model grid cell can be advected farther downstream in a different grid cell before being dissipated (Nakanishi and Niino, 2006; Krishnamurthy et al., 2011; Hong and Dudhia, 2012).

The inaccuracy of the mesoscale model representation of ϵ impacts a wide variety of processes that are controlled by the TKE dissipation rate. In fact, the dissipation of turbulence affects the development and propagation of forest fires (Coen et al., 2013), it has consequences on aviation meteorology and potential aviation accidents (Gerz et al., 2005; Thobois et al., 2015), it regulates the dispersion of pollutants in the boundary layer (Huang et al., 2013), and it affects wind energy applications (Kelley et al., 2006); for example, in terms of the development and erosion of wind turbine wakes (Bodini et al., 2017).

Several studies have documented the variability of ϵ using observations from both in-situ (Champagne et al., 1977; Oncley et al., 1996; Frehlich et al., 2006) and remote-sensing instruments (Frehlich, 1994; Smalikho, 1995; Shaw and LeMone, 2003). Bodini et al. (2018, 2019b) showed how ϵ has strong diurnal and annual cycles onshore, with topography playing a key role in triggering its variability. On the other hand, offshore turbulence regimes (Bodini et al., 2019a) are characterized by smaller values of ϵ, with cycles mostly impacted by wind regimes rather than convective effects. Also, ϵ greatly increases in the wakes of obstacles, for example wind turbines (Lundquist and Bariteau, 2015; Wildmann et al., 2019) or whole wind farms (Bodini et al., 2019b).

This knowledge on the variability of TKE dissipation rate provided by observations lays the foundation to explore innovative ways to improve the model representation of ϵ in the atmospheric boundary layer. In this study, we leverage the potential of machine-learning techiniques to explore their potential application to improve the parameterizations of ϵ. Machine-learning techniques can successfully capture the complex and nonlinear relationship between multiple variables without the need of representing the physical process that governs this relationship. They have been successfully used to advance the understanding of several atmospheric processes, such as convection (Gentine et al., 2018), turbulent fluxes (Leufen and Schädler, 2018), and precipitation nowcasting (Xingjian et al., 2015). The renewable energy sector has also experienced various applications of machine-learning techniques, in both solar (Sharma et al., 2011; Cervone et al., 2017) and wind (Giebel et al., 2011; Optis and Perr-Sauer, 2019) power forecasting. Applications have also been explored at the wind turbine level, for turbine power curve modeling (Clifton et al., 2013), turbine faults and controls (Leahy et al., 2016), and turbine blade management (Arcos Jiménez et al., 2018).

Here, we train and test different machine-learning algorithms to predict ϵ from a set of atmospheric and topographic variables. Section 2 describes the Perdigão field campaign and how we retrieved ϵ from the sonic anemometers on the meteorological towers. In Section 3, we then evaluate the accuracy of one of the most common planetary boundary layer parameterization schemes used in numerical weather prediction: the Mellor, Yamada, Nakanishi, and Niino (MYNN) parameterization scheme

https://doi.org/10.5194/gmd-2020-16
Preprint. Discussion started: 21 April 2020

Geoscientific Model Development Discussions
Open Access
EGU

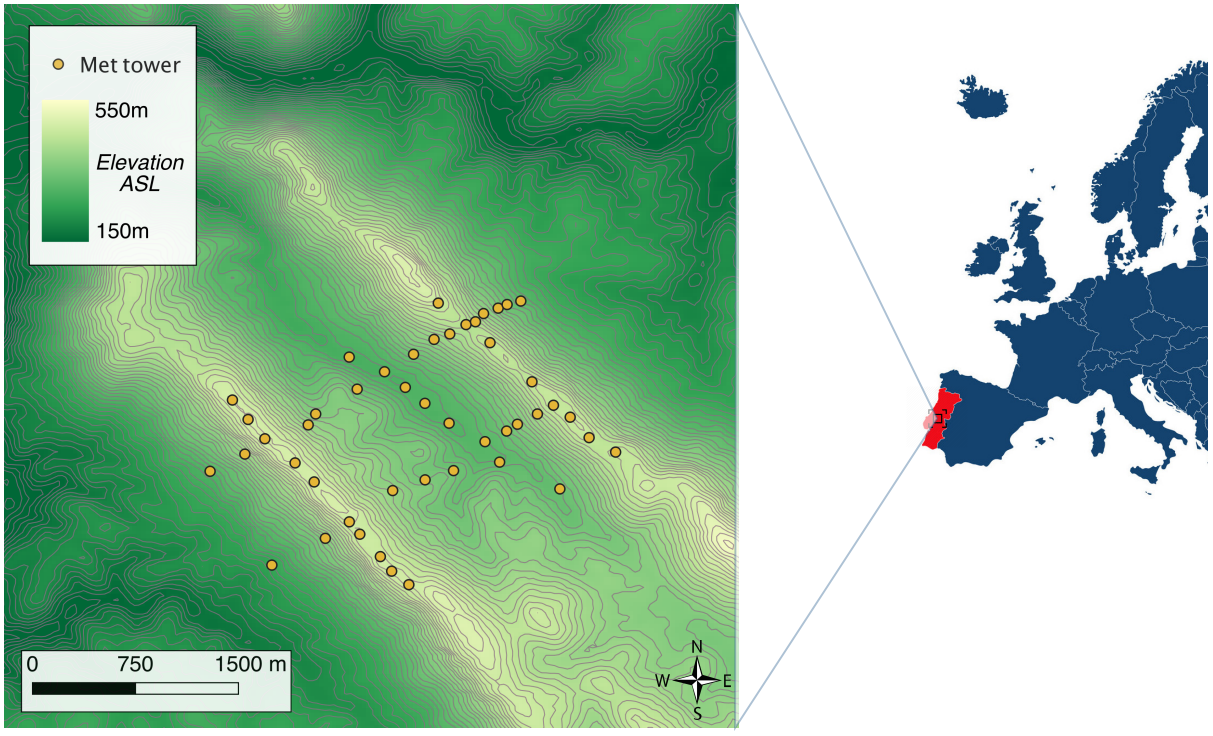

Figure 1. Map of the Perdigão valley showing the location of the 48 meteorological towers whose data are used in this study.

(Nakanish, 2001). Section 4 presents the machine-learning algorithms that we used in our analysis. The results of our study are shown in Section 5, and discussed in Section 6, where future work is suggested.

2 Data

2.1 Meteorological towers at the Perdigão field campaign

The Perdigão field campaign (Fernando et al., 2018), an international cooperation between several universities and research institutes, brought an impressive number of instruments to a valley in central Portugal to survey the atmospheric boundary layer in complex terrain. The Perdigão valley is limited by two mountain ridges running from northwest to southeast (Figure 1), separated by ∼1.5 km. The intensive operation period (IOP) of the campaign, used for this study, was from 1 May to 15 June 2017.

At Perdigão, 184 sonic anemometers were mounted on 48 meteorological towers, which provided an unprecedented density of instruments in such a limited domain (Figure 1). Observations from the sonic anemometers (a mix of Campbell Scientific CSAT3, METEK uSonic, Gill WindMaster, and YOUNG Model 81000 instruments) are recorded at a 20-Hz frequency. The height of the towers ranged from 2 m to 100 m, with the sonic anemometers mounted at various heights (histogram in Figure

https://doi.org/10.5194/gmd-2020-16
Preprint. Discussion started: 21 April 2020

Geoscientific Model Development Discussions

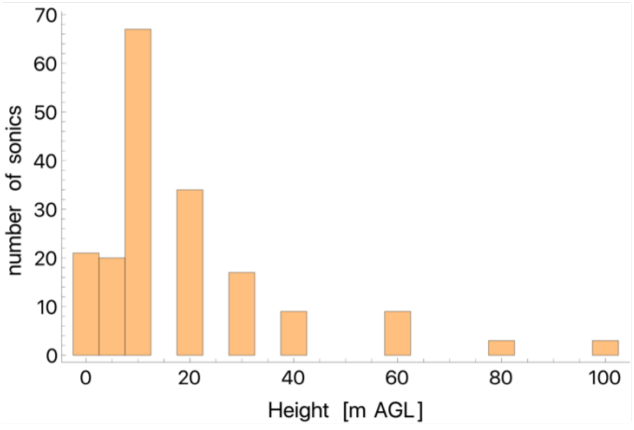

Figure 2. Histogram of the heights of the 184 sonic anemometers considered in this analysis. The bin for the lowest height includes sonic anemometers at 2 m above ground level (AGL).

2), allowing for an extensive survey of the variability of the wind flow in the boundary layer. Data from the sonic anemometers have been tilt-corrected following the planar fit method (Wilczak et al., 2001).

2.2 TKE dissipation rate from sonic anemometers

TKE dissipation rate from the sonic anemometers on the meteorological towers is calculated from the second-order structure function of the horizontal velocity (Muñoz-Esparza et al., 2018):

$$\epsilon = \frac{1}{U\tau}\left[aD_U(\tau)\right]^{3/2} \tag{1}$$

where $D_U(\tau)$ is the second-order structure function of the horizontal velocity, U, calculated over temporal increments, τ, and $a = 0.52$ is the one-dimensional Kolmogorov constant (Paquin and Pond, 1971; Sreenivasan, 1995). ϵ is calculated every 30 s, and the fit to the Kolmogorov model (Kolmogorov, 1941; Frisch, 1995) is done using a temporal separation between $\tau_1 = 0.1$ s and τ_2 = 2 s (Bodini et al., 2018). To exclude tower wake effects, data have been discarded when the recorded wind direction was within $\pm$ 30° of the direction of the tower boom. Data during precipitation periods (as recorded by a precipitation sensor on the tower 'riSW06' on the southwest ridge) have also been discarded from further analysis.

2.3 Input features for machine-learning algorithms

In order to train machine-learning algorithms to predict (the logarithm of) ϵ, a set of input features needs to be provided. By taking advantage of the main findings of the observational studies on the variability of ϵ, we decided to use both atmospheric- and terrain-related variables to capture the impact of topography on atmospheric turbulence. Therefore, we use the following input features (calculated at the same location and height as ϵ) for all of the considered algorithms:

- wind speed;

https://doi.org/10.5194/gmd-2020-16
Preprint. Discussion started: 21 April 2020


Geoscientific Model Development Discussions

– TKE, calculated as

$$TKE = \frac{1}{2}\left(\sigma_u^2 + \sigma_v^2 + \sigma_w^2\right) \tag{2}$$

where the variances of the wind components are calculated over 2-minute intervals;

– $\mathrm{TKE}^{3/2}$, which is the main variable currently used to represent ϵ in parameterizations in numerical weather prediction models (Nakanish, 2001);

– friction velocity, calculated as

$$u_* = (\overline{u'w'}^2 + \overline{v'w'}^2)^{1/4}. \tag{3}$$

An averaging period of 30 minutes (De Franceschi and Zardi, 2003; Babić et al., 2012) has been used to apply the Reynolds decomposition and calculate average quantities and fluctuations.

– height above the ground z_{son};

– Obukhov length, L, calculated from each sonic anemometer to classify atmospheric stability. We calculate L as

$$L = -\frac{\overline{\theta_v} \cdot u_*^3}{k \cdot g \cdot \overline{w'\theta_v'}}. \tag{4}$$

θ_v is the virtual potential temperature (K); $u_* = (\overline{u'w'}^2 + \overline{v'w'}^2)^{1/4}$ is the friction velocity ($\mathrm{m\,s^{-1}}$); $k = 0.4$ is the von Kármán constant; $g = 9.81\,\mathrm{m\,s^{-2}}$ is the gravity acceleration; and $\overline{w'\theta_v'}$ is the kinematic sensible heat flux ($\mathrm{m\,K\,s^{-1}}$). For atmospheric stability, we classify neutral conditions as $L \leq$ -500 m and $L > 500$ m; unstable conditions as -500 m $< L \leq 0$ m; and stable conditions as 0 m $< L \leq 500$ m (Muñoz-Esparza et al., 2012);

– the ratio z_{son}/L;

– the standard deviation std(z_{terr}) of the terrain elevation in a 1-km radius sector centered on the measurement point (i.e., the location of the sonic anemometer). The angular extension of the sector is set equal to $\pm 30°$ from the recorded wind direction (an example is shown in Figure 3). In this way, we can capture the influence of upwind topography to trigger turbulence.

The distribution of the input features and of $\log(\epsilon)$ are shown in the Supplement.

While we acknowledge that the input features are not fully uncorrelated, we found that including all these features provides a better predictive power for the learning algorithms, despite negatively affecting the computational requirements of the training phase. The application of principal component analysis can help reduce the number of dimensions in the input features while preserving the predictive power of each, but it is beyond the scope of the current work.

https://doi.org/10.5194/gmd-2020-16
Preprint. Discussion started: 21 April 2020

Geoscientific Model Development Discussions
Open Access
EGU

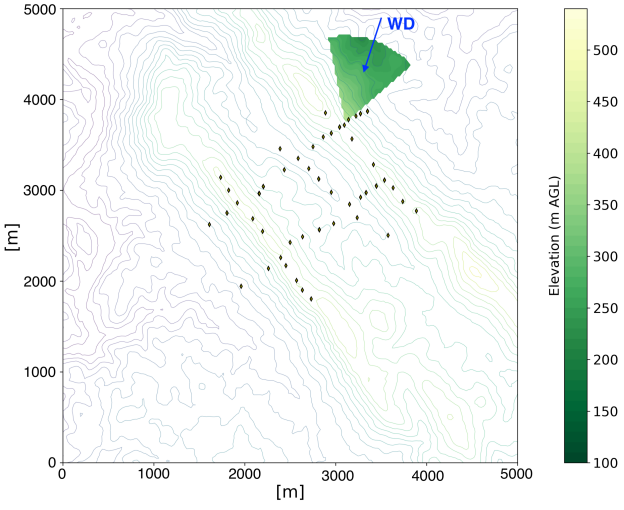

Figure 3. Example of an upwind terrain elevation sector with a 1-km radius centered on the location of one of the meteorological towers at Perdigão.

3 Accuracy of current parameterization of TKE dissipation rate in mesoscale models

Before testing the performance of machine-learning algorithms in predicting TKE dissipation rates, we first assess the current accuracy of the parameterization of ϵ in numerical models. In the Weather Research and Forecasting model (WRF) (Skamarock et al., 2005), the most widely used numerical weather prediction model, turbulence in the boundary layer can be represented with several planetary-boundary-layer (PBL) schemes, most of which implicitly assume a local balance between turbulence production and dissipation. Among the different PBL schemes, the MYNN scheme is one of the most commonly chosen. Turbulence dissipation rate in MYNN is given as a function of TKE as (Nakanish, 2001)

$$\epsilon = \frac{(2\,\mathrm{TKE})^{3/2}}{B_1\,L_M} \tag{5}$$

where $B_1 = 24$, and the master length scale, L_M, is defined with a diagnostic equation, based on large-eddy simulations, as a function of three other length scales

$$\frac{1}{L_M} = \frac{1}{L_S} + \frac{1}{L_T} + \frac{1}{L_B}. \tag{6}$$

L_S is the length scale in the surface layer, given by

$$L_S = \begin{cases} \kappa\, z/3.7 & \zeta \geq 1 \\ \kappa\, z\,(1+2.7\,\zeta)^{-1} & 0 \leq \zeta < 1 \\ \kappa\, z\,(1-\alpha_4\,\zeta)^{0.2} & \zeta < 0 \end{cases} \tag{7}$$

https://doi.org/10.5194/gmd-2020-16
Preprint. Discussion started: 21 April 2020


Geoscientific Model Development Discussions

where $\kappa = 0.4$ is the von Kármán constant, $\zeta = z/L$ (with L the Obukhov length), $\alpha_4 = 100.0$.

L_T is the length scale depending upon the turbulent structure of the PBL (Mellor and Yamada, 1974), defined as

$$L_T = \alpha_1 \frac{\int_0^\infty q\ z\ dz}{\int_0^\infty q\ dz} \tag{8}$$

where $q = \sqrt{2\,TKE}$, and $\alpha_1 = 0.23$.

L_B is a length scale limited by the buoyancy effect, given by

$$L_B = \begin{cases} \alpha_2\ q/N & \partial\Theta/\partial z > 0 \text{ and } \zeta \geq 0 \\ \frac{\alpha_2\ q + \alpha_3\ q\ (q_c/L_T\ N)^{1/2}}{N} & \partial\Theta/\partial z > 0 \text{ and } \zeta < 0 \\ \infty & \partial\Theta/\partial z \leq 0 \end{cases} \tag{9}$$

with N the Brunt-Väisälä frequency, Θ the mean potential temperature, $\alpha_2 = 1.0$, and $\alpha_3 = 5.0$.

From the available observations from the meteorological towers at Perdigão, only L_S can be determined, while the calculation of L_T and L_B would only be possible with critical assumptions about the vertical profile of TKE. Therefore, we decide to approximate L_M as

$$\frac{1}{L_M} \approx \frac{1}{L_S}. \tag{10}$$

By doing so, L_M is overestimated, which in turn implies that ϵ as calculated from Eq. (5) will be underestimated.

To evaluate the accuracy of the MYNN parameterization of ϵ, we calculated, every 30 s, the parameterized ϵ using Eq. (5) (with the approximation in Eq. (10)) from all of the 184 sonic anemometers considered in the study, and compared with the observed values of TKE dissipation rate (Figure 4) derived from the sonic anemometers with Eq. (1). Given the extremely large range of variability of ϵ, we calculate all the error metrics using the logarithm of predicted and observed ϵ. The TKE dissipation rate predicted by the MYNN parameterization shows, on average, a large positive bias compared to the observed values, with a mean bias (in logarithmic space) of +28%. This bias would be even larger if L_M was calculated including all the contributions according to Eq. (6), and not L_s only as in our approximation. The root-mean-square error (RMSE) is 0.76, and the mean absolute error (MAE) is 0.57.

Different atmospheric stability conditions give different biases. Figure 5 compares observed and parameterized ϵ values for stable and unstable conditions, classified based on the Obukhov length, L, measured at each sonic anemometer. Stable cases show the largest bias (mean of +63% in terms of the logarithms of ϵ), whereas for unstable conditions the bias is smaller (mean of +7% in terms of the logarithms of ϵ). The MYNN parameterization of ϵ is therefore especially inadequate to represent small values of ϵ, which mainly occur in stable conditions.

Different heights also impact the accuracy of the parameterization of ϵ. As shown in Figure 6, the mean bias in parameterized $\log(\epsilon)$ decreases with height. Close to the surface (data from the sonic anemometers at 2 m AGL), a mean bias (in logarithmic space) of almost +30% is found, whereas for the sonic anemometers at 100 m AGL, we find a mean bias of $\sim$ +20%. This difference in bias with height becomes much larger if the bias is calculated on the actual ϵ values (and not on their logarithm).

https://doi.org/10.5194/gmd-2020-16
Preprint. Discussion started: 21 April 2020

Geoscientific Model Development Discussions
Open Access
EGU

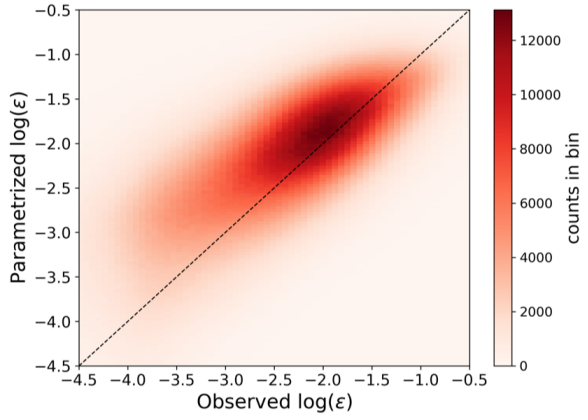

Figure 4. Density histogram showing the comparison between observed and MYNN-parameterized ϵ from the 184 sonic anemometers at Perdigão.

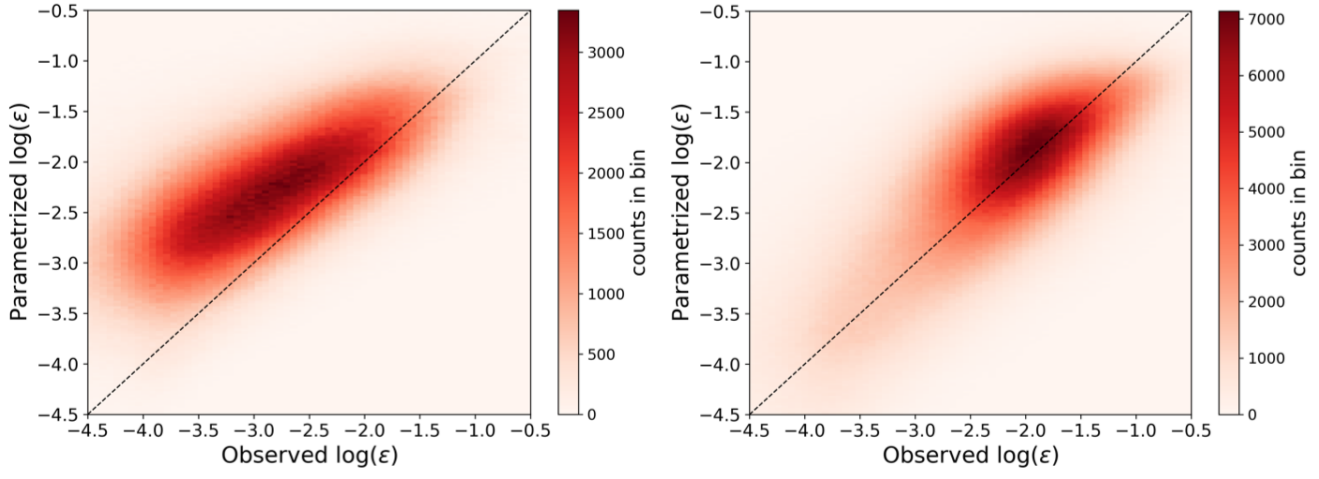

Figure 5. Density histogram showing the comparison between observed and MYNN-parameterized ϵ from the 184 sonic anemometers at Perdigão for stable conditions (left) and unstable conditions (right), as quantified by the Obukhov length calculated at each sonic anemometer.

Therefore, our results show how the MYNN formulation fails in accurately representing atmospheric turbulence especially in the lowest part of the boundary layer.

https://doi.org/10.5194/gmd-2020-16
Preprint. Discussion started: 21 April 2020

Geoscientific Model Development Discussions
Open Access
EGU

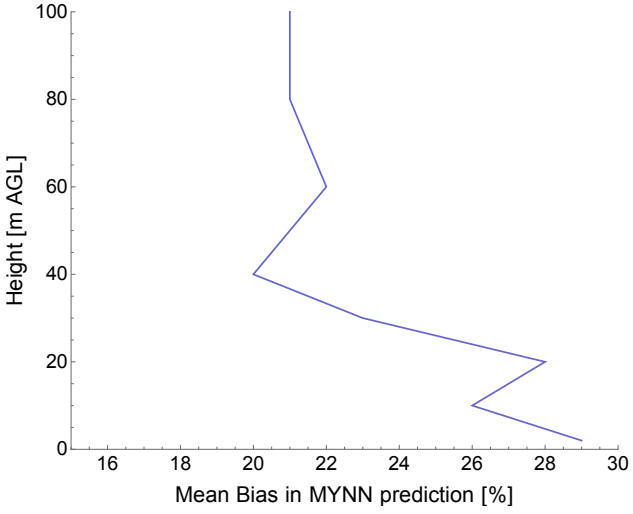

Figure 6. Mean bias in the MYNN-parameterized $\log(\epsilon)$ at different heights, as calculated from the 184 sonic anemometers at Perdigão.

4 Machine-learning algorithms

To test the power of machine learning to improve the numerical representation of the TKE dissipation rate, we consider three learning algorithms in this study: multivariate linear regression, multivariate third-order polynomial regression, and random forest. These models are trained and tested to predict the logarithm of ϵ. For all but the random forest algorithm, the data were scaled and normalized. Time stamps with missing data were removed from the analysis.

For the purpose of machine-learning algorithms, the data set has to be divided into three subsets: training, validation, and testing sets (Friedman et al., 2001). The algorithms are first trained multiple times with different hyperparameters (model parameters set before the training phase) on the training set, then the validation set is used to choose the best set of hyperparameters, and finally the predicting performance of the trained algorithm is assessed on the testing set. Usually, the data set is split randomly into training, validation, and testing sets. However, as the data used in this study consist of observations taken every 30 s, data in contiguous time stamps are likely characterized by a large auto-correlation. Therefore, the traditional random split between training and testing data would lead to an artificially enhanced performance of the machine-learning algorithms, which would be tested on data with a large auto-correlation with the ones used for the training. Therefore, here we use one concurrent week of the data for testing ($\sim$17% of the data), whereas the other 5 weeks are split between training (4 weeks, 66% of the data) and validation (1 week, 17% of the data). The 1-week testing period is shifted continuously throughout the considered 6 weeks of observations at Perdigão, so that each model is trained and its prediction performance tested six times. For each algorithm, we evaluate the overall performance based on the RMSE between the actual and predicted (logarithm of) ϵ, averaged over the different week-long testing periods.

Before testing the models, however, it is important to avoid overfitting by setting the values of hyperparameters. To test different combinations of hyperparameters and determine the best set, we use cross validation with randomized search, with

https://doi.org/10.5194/gmd-2020-16
Preprint. Discussion started: 21 April 2020

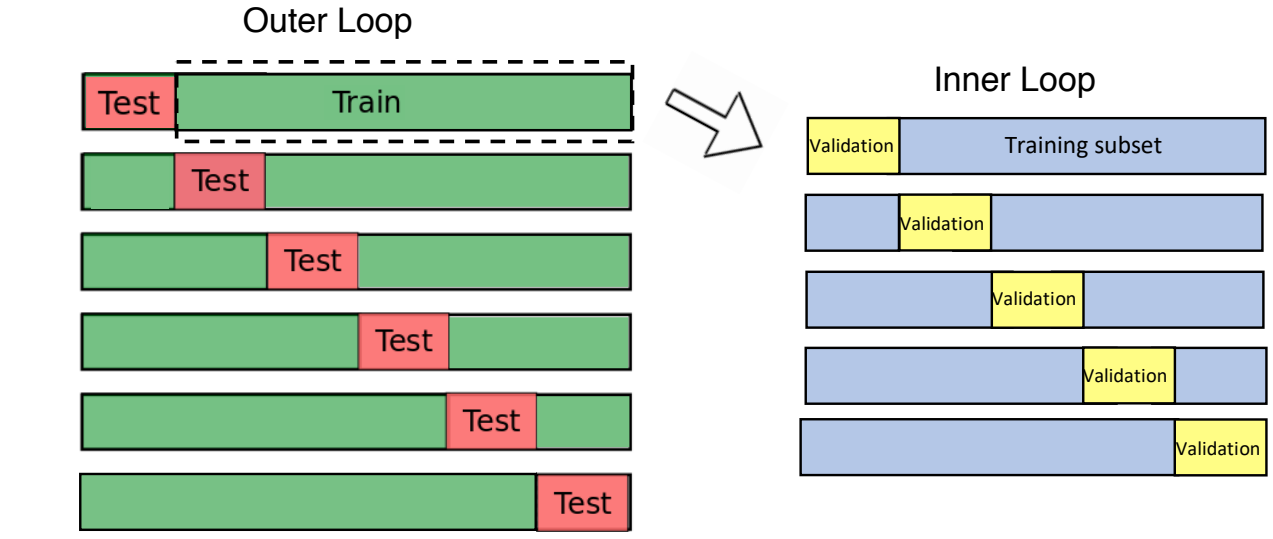

Figure 7. Cross-validation approach used to evaluate the performance of the machine-learning models considered in this study.

five parameter settings sampled for each learning algorithm. For each set of hyperparameters, the RMSE between the actual and predicted $\log(\epsilon)$ in the validation test is calculated. For each model, we select the hyperparameter combination (among the ones surveyed in the cross validation) that leads to the lowest mean (across the five validation sets) RMSE. We then use this set as the final combination for assessing the performance of the models on the testing set. Overall, the procedure is repeated six times, by shifting the 1-week testing set (Figure 7).

In the following paragraphs, we describe the main characteristics of the three machine-learning algorithms used in our study. A more detailed description can be found in machine-learning textbooks (Géron, 2017).

4.1 Multivariate linear regression

To check whether simple learning algorithms can improve the current numerical parameterization of ϵ, we test the accuracy of multivariate linear regression

$$\log(\hat{\epsilon}) = \theta_0 + \theta_1\, x_1 + \theta_2\, x_2 + ... + \theta_n\, x_n \tag{11}$$

where $\hat{\epsilon}$ is the machine-learning predicted value of ϵ, n is the number of features used to predict ϵ (here 8), x_i is the i^{th} feature value, and θ_j is the j^{th} model parameter.

To avoid training a model that overfits the data, regularization techniques need to be implemented, so that the learning model is constrained: the fewer degrees of freedom the model has, the harder it will be for it to overfit the data. We use Ridge regression (Hoerl and Kennard, 1970) (`Ridge` in Scikit-learn) to constrain the multivariate regression. Ridge regression constrains the weights of the model to have them stay as small as possible. The Ridge regression is achieved by adding a regularization term

https://doi.org/10.5194/gmd-2020-16
Preprint. Discussion started: 21 April 2020

Geoscientific Model Development Discussions
Open Access
EGU

to the cost function (MSE)

$$J(\theta) = MSE(\theta) + \alpha \sum_{i=1}^{n} \theta_i^2 \tag{12}$$

where the hyperparameter, α, controls how much the model will be regularized. The optimal value of the hyperparameter, α, is determined by cross validation, as described earlier, with values sampled in the range from 0.1–10.

4.2 Multivariate third-order polynomial regression

Multivariate polynomial regression can easily be achieved by adding powers of each input feature as new features. The regression algorithm is then trained as a linear model on this extended set of features. For a third-order polynomial regression, the model becomes

$$\begin{aligned}\log(\hat{\epsilon}) = \theta_0 &+ \sum_{i=1}^{n} \theta_i \, x_i + \sum_{i=1}^{n} \theta_{ii} \, x_i^2 + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \theta_{ij} \, x_i \, x_j \\ &+ \sum_{i=1}^{n} \theta_{iii} \, x_i^3 + \sum_{i=1}^{n} \sum_{j \neq i} \theta_{iij} \, x_i^2 \, x_j \\ &+ \sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} \sum_{k=j+1}^{n-1} \theta_{ijk} \, x_i \, x_j \, x_k \end{aligned} \tag{13}$$

Ridge regression (`Ridge` in Scikit-learn) is used again to constrain the multivariate polynomial regression, with the hyperparameter α in Eq. (12) determined via cross validation, with values sampled in the range from 1–2000.

4.3 Random forest

Random forests (`RandomForestRegressor` in Scikit-learn) combine multiple decision trees to provide an ensemble prediction.

A decision tree can learn patterns and then predict values by recursively splitting the training data based on thresholds of the different input features.

As a result, the data are divided into groups, each associated with a single predicted value of ϵ, calculated as the average target value (of the observed ϵ) of the instances in that group.

As an ensemble of decision trees, a random forest trains them on different random subsets of the training set. Once all the predictors are trained, the ensemble (i.e., the random forest) can make a prediction for a new instance by taking the average of all the predictions from the single trees. In addition, random forests introduce some extra randomness when growing trees: instead of looking for the feature that, when split, reduces the overall error the most when splitting a node, a random forest searches for the best feature among a random subset of features.

Decision trees make very few assumptions about the training data. As such, if unconstrained, they will adapt their structure to the training data, fitting them closely, and most likely overfitting them, without then being able to provide accurate predictions on new data. To avoid overfitting, regularization can be achieved by setting various hyperparameters that insert limits to the

https://doi.org/10.5194/gmd-2020-16
Preprint. Discussion started: 21 April 2020

Geoscientific Model Development Discussions

Table 1. Hyperparameters considered for the random forest algorithm.

Hyperparameter	Meaning	Sampled values
Number of estimators	Number of trees in the forest	10 - 250
Maximum depth	Maximum depth of the tree	1 - 50
Maximum number of leaf nodes	Maximum number of leaf nodes in the decision tree	2000 - 500,000
Maximum number of features	Number of features to consider when looking for the best split	1 - 8
Minimum number of samples to split	Minimum number of samples required to split an internal node	1 - 200
Minimum number of samples for a leaf	Minimum number of samples required to be at a leaf node	1 - 50

structure of the trees used to create the random forests. Table 1 shows the hyperparameters considered here and the values we sampled in the cross-validation search (with five sets of parameters sampled).

5 Results

To evaluate the prediction performance of the three machine-learning algorithms we considered, we use, for each method, a density histogram showing the comparison between observed and machine-learning-predicted ϵ. As error metrics, we compare RMSE, MAE, and coefficient of determination (R^2) of the machine-learning predictions with what we obtained from the MYNN parameterization, with the caveat that while the MYNN scheme is thought to provide a universal representation of ϵ, the machine-learning models have been specifically trained on data from a single field campaign.

Each machine-learning algorithm was tested on six 1-week-long testing periods, as described in Section 4. For each method we present the RMSE, MAE, and R^2 averaged across the different testing periods. Because the length scale approximation we made in calculating MYNN-predicted ϵ lead to an optimistic result, the RMSE and MAE for the MYNN case would in reality be higher, and the R^2 lower, and so the error reductions achieved with the machine-learning algorithms would even be greater than the numbers we report here.

Predictions from the multivariate linear regression (Figure 8) only show a slight improvement constrained to values of ϵ greater than 10^{-3} m^2 s^{-3}, whereas a severe overestimation of ϵ is found for low turbulence regimes, when the linear regression predicts almost constant values of ϵ. We speculate that this limited performance is mainly because of the fact that the linear regression can only accurately model ϵ over the same range of orders of magnitude of TKE (which is the dominant input feature, as will be systematically proven with the random forest algorithm), while the dissipation rate varies over more orders of magnitude than TKE (see histograms in the Supplement). Despite the poor agreement with the observations, the multivari-

https://doi.org/10.5194/gmd-2020-16
Preprint. Discussion started: 21 April 2020

Geoscientific Model Development Discussions
Open Access
EGU

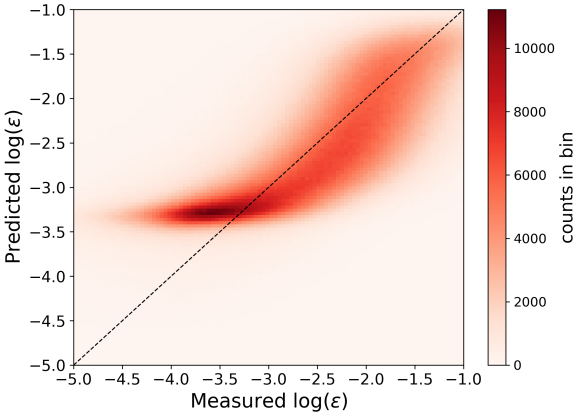

Figure 8. Density histogram showing the comparison between observed and machine-learning-predicted ϵ from a multivariate linear regression on the test set.

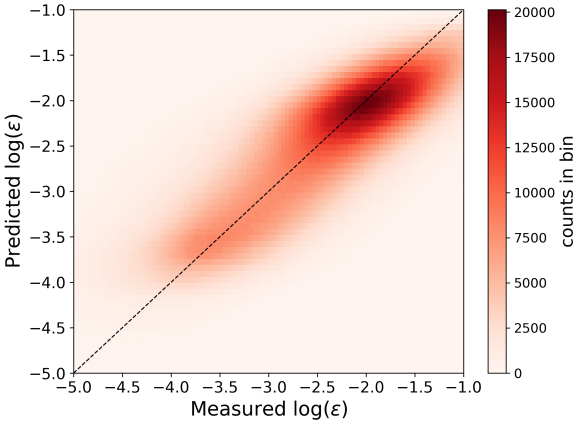

Figure 9. Density histogram showing the comparison between observed and machine-learning-predicted ϵ from a multivariate third-order polynomial regression on the test set.

ate linear regression improves, on average, on MYNN. Overall, the average RMSE (0.66) is 13% smaller than the MYNN parameterization, and the average MAE (0.51) is 12% lower than the MYNN prediction.

On the other hand, the multivariate third-order polynomial regression provides a significant improvement (Figure 9) for the representation of ϵ, by eliminating the average bias found in the MNYNN parameterization. The additional input features created by the polynomial model allow for an accurate prediction of ϵ even at the low turbulence regime, as they allow the model to cover the full range of variability of ϵ. We find the average RMSE (0.59) to be over 23% smaller than the MYNN parameterization, and the average MAE (0.45) is 22% lower than the MYNN representation.

https://doi.org/10.5194/gmd-2020-16
Preprint. Discussion started: 21 April 2020

Geoscientific Model Development Discussions
Open Access
EGU

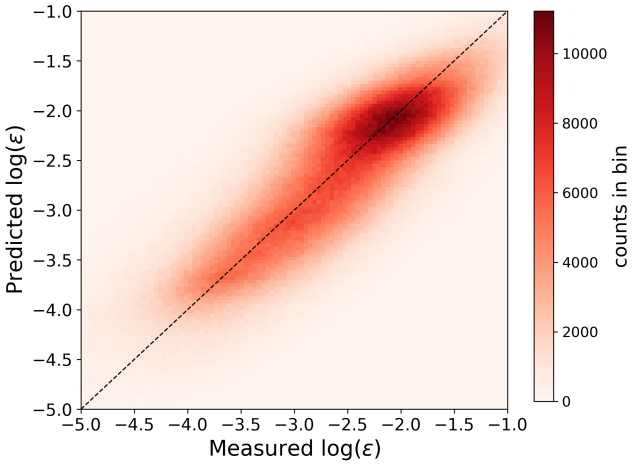

Figure 10. Density histogram showing the comparison between observed and machine-learning-predicted ϵ from a random forest on the test set.

Results from the random forest (Figure 10) provides accurate predictions at both high and low turbulence regimes. The RMSE (0.53) is reduced by about 30% from the MYNN case, and the MAE (0.41) by 29%, with no average bias between observed and predicted values of ϵ.

Table 2 summarizes the performance of all of the considered algorithms, and includes results also in terms of the coefficient of determination, R^2.

Table 2. Performance of the machine-learning algorithms trained and tested at Perdigão, measured in terms of RMSE, MAE, and R^2 between the logarithm of observed and MYYN-parameterized ϵ.

	MYNN parameterization	Linear regression	Third-order polynomial regression	Random forest
RMSE	0.76	0.66	0.59	0.53
% change in RMSE		-13%	-23%	-30%
MAE	0.58	0.51	0.45	0.41
% change in MAE		-12%	-22%	-29%
R^2	0.40	0.53	0.62	0.68
% change in R^2		+33%	+55%	+70%

https://doi.org/10.5194/gmd-2020-16
Preprint. Discussion started: 21 April 2020

Geoscientific Model Development Discussions

Table 3. Feature importance classification as derived from the random forest.

Input feature	Feature importance
TKE	33%
$\text{TKE}^{3/2}$	29%
z_{son}	11%
Wind Speed	8%
u_{star}	8%
std(z_{terr})	6%
z_{son}/L	3%
L	2%

5.1 Feature importance

Random forests allow for an assessment of the relative importance of the input features used to predict (the logarithm of) ϵ. The importance of a feature is calculated by looking at how much the tree nodes that use that feature reduce the MSE on average (across all trees in the forest), weighted by the number of times the feature is selected. Table 3 shows the feature importance for the eight input features we used in this study.

The feature importance results are affected by the correlation between some of the input features used in the models. We find how turbulence kinetic energy is the preferred feature for tree splitting, with the largest importance (62% in total, considering TKE and $\text{TKE}^{3/2}$) in reducing the prediction error for ϵ in the random forest. This result, which can be expected as both TKE and ϵ are variables connected to turbulence in the boundary layer, agrees well with the current formulation of the MYNN parameterization of ϵ, which includes TKE as main term. As TKE is correlated to u_{star} and L, we find that the decision trees more often split the data based on TKE, so that the feature importance of its correlated variables is found to be much lower. The limitations of the Monin-Obukhov similarity theory (Monin and Obukhov, 1954) in complex terrain might also be an additional cause for the low feature importance of the features associated with L. The height of the measurement is the next most important feature (11%), followed by wind speed (at 8%). The standard deviation of the upwind elevation has an importance of 6%. Though not negligible, the importance of topography might increase by considering different parameters that could better capture the topographic effect.

5.2 Partial dependence plots

To assess the dependence of TKE dissipation rate on the individual features considered in this study, Figure 11 shows partial dependence plots for the features with the largest importance, namely wind speed, friction velocity, TKE, and z (plots for the rest of the features are included in the Supplement). These are obtained, for each input feature, by applying the machine-learning algorithm (here, random forests using about 120,000 instances to speed up the computation) multiple times with the

https://doi.org/10.5194/gmd-2020-16
Preprint. Discussion started: 21 April 2020

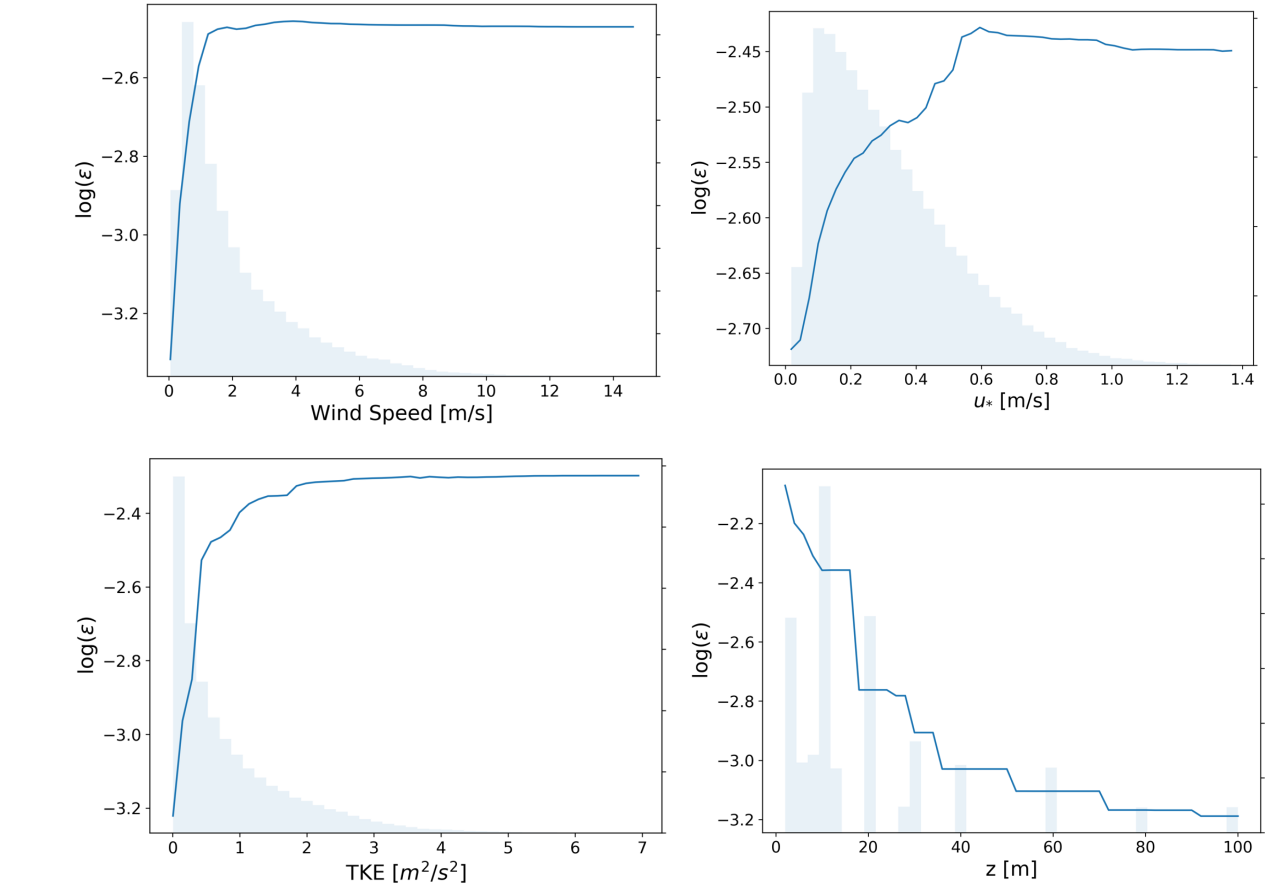

Figure 11. Partial dependence plots for wind speed; friction velocity, u_*; TKE; and measurement height, z. Distributions of the considered features are shown in the background.

other feature variables constant (at their means) while varying the target input feature and measuring the effect on the response variable (here, $\log(\epsilon)$).

A strong relationship appears between TKE dissipation and low wind speed, with a steep increase in ϵ for wind speeds up to $\sim$2 m s^{-1}. Stronger wind speeds do not appear to have a systematic impact on ϵ. The relationship between TKE and its dissipation shows a similar trend, with ϵ systematically increasing for TKE values up to $\sim$2 m^2 s^{-2}, and staying constant for stronger TKE values. TKE dissipation also increases with friction velocity, u_*, with a less steep trend. Finally, height above the ground shows a systematic impact on ϵ, with a strong downward trend.

https://doi.org/10.5194/gmd-2020-16
Preprint. Discussion started: 21 April 2020


Geoscientific Model Development Discussions
Open Access
EGU

6 Conclusions

Despite turbulence being a fundamental quantity for the development of multiple phenomena in the atmospheric boundary layer, the current representations of TKE dissipation rate (ϵ) in numerical weather prediction models suffer from large inaccuracies. In this study, we quantified the error introduced in the MYNN parameterization of ϵ by comparing predicted and observed values of ϵ from 184 sonic anemometers from 6 weeks of observations at the Perdigão field campaign. A large positive bias (average +28% in logarithmic space) emerges, with larger errors found in atmospheric stable conditions. The need for a more accurate representation of ϵ is therefore clearly demonstrated.

The results of this study show how machine learning can provide new ways to successfully represent TKE dissipation rate from a set of atmospheric and topographic parameters. While ultrasimple models such as linear regression cannot provide a satisfactory representation of ϵ, more sophisticated algorithms, such as a random forest approach, lead to over a 30% reduction in the average error introduced in the parameterization of ϵ and eliminate the large bias found in it, for the Perdigão field campaign. Although the generalization gap between the universal nature of the MYNN parameterization of ϵ and the campaign-specific training and testing of the machine-learning models has to be acknowledged, the results of this study can be considered as a proof of concept of the potentialities of machine-learning-based representations of complex atmospheric processes.

Multiple opportunities exist to extend the work presented here. In the future, additional learning algorithms, such as support vector machines and extremely randomized trees, should be considered. Deep learning methods, such as recurrent neural networks, and specifically long-short term memory, which are well-suited for time-series-based problems, could also be considered to obtain a more complete overview of the capabilities of machine-learning techniques for improving numerical representations of ϵ. Moreover, additional input features could be added to the learning algorithms to possibly identify additional variables with a large impact on atmospheric turbulence. To this extent, we expect vegetation height to be an important parameter to consider. The results could also be analyzed with a focus on the temporal variability of the performance of the proposed algorithms; for example, at different times of the day, or for different atmospheric stability conditions. Finally, the learning algorithms developed here would need to be tested using data from different field experiments, to understand whether the results obtained in this study can be generalized everywhere. Once the performance of a machine-learning representation of ϵ has been accurately tested, its implementation in numerical weather prediction models, such as the Weather Research and Forecasting model, should be achieved.

Code and data availability. High-resolution data from sonic anemometers on the meteorological towers (UCAR/NCAR, 2019) are available through the EOL project at https://doi.org/10.26023/8X1N-TCT4-P50X. The random forest code used for the analysis is stored at https://doi.org/10.5281/zenodo.3754710.

Author contributions. NB analyzed the data from the sonic anemometers and applied the machine-learning figures, in close consultation with JKL and MO. NB wrote the paper, with significant contributions from JKL and MO.

https://doi.org/10.5194/gmd-2020-16
Preprint. Discussion started: 21 April 2020

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. Support to NB and JKL is provided by the National Science Foundation, under the CAREER program AGS-1554055 and the award AGS-1565498. This work utilized the RMACC Summit supercomputer, which is supported by the National Science Foundation (awards ACI-1532235 and ACI-1532236), the University of Colorado Boulder, and Colorado State University. The Summit supercomputer is a joint effort of the University of Colorado Boulder and Colorado State University.

https://doi.org/10.5194/gmd-2020-16
Preprint. Discussion started: 21 April 2020

Geoscientific Model Development Discussions

References

Albertson, J. D., Parlange, M. B., Kiely, G., and Eichinger, W. E.: The average dissipation rate of turbulent kinetic energy in the neutral and unstable atmospheric surface layer, Journal of Geophysical Research: Atmospheres, 102, 13 423–13 432, 1997.

Arcos Jiménez, A., Gómez Muñoz, C., and García Márquez, F.: Machine learning for wind turbine blades maintenance management, Energies, 11, 13, 2018.

Babić, K., Bencetić Klaić, Z., and Večenaj, Ž.: Determining a turbulence averaging time scale by Fourier analysis for the nocturnal boundary layer, Geofizika, 29, 35–51, 2012.

Berg, L. K., Liu, Y., Yang, B., Qian, Y., Olson, J., Pekour, M., Ma, P.-L., and Hou, Z.: Sensitivity of Turbine-Height Wind Speeds to Parameters in the Planetary Boundary-Layer Parametrization Used in the Weather Research and Forecasting Model: Extension to Wintertime Conditions, Boundary-Layer Meteorology, pp. 1–12, 2018.

Bodini, N., Zardi, D., and Lundquist, J. K.: Three-dimensional structure of wind turbine wakes as measured by scanning lidar, Atmospheric Measurement Techniques, 10, 2017.

Bodini, N., Lundquist, J. K., and Newsom, R. K.: Estimation of turbulence dissipation rate and its variability from sonic anemometer and wind Doppler lidar during the XPIA field campaign, Atmospheric Measurement Techniques, 11, 4291–4308, 2018.

Bodini, N., Lundquist, J. K., and Kirincich, A.: US East Coast Lidar Measurements Show Offshore Wind Turbines Will Encounter Very Low Atmospheric Turbulence, Geophysical Research Letters, 2019a.

Bodini, N., Lundquist, J. K., Krishnamurthy, R., Pekour, M., Berg, L. K., and Choukulkar, A.: Spatial and temporal variability of turbulence dissipation rate in complex terrain, Atmospheric Chemistry and Physics, 19, 4367–4382, 2019b.

Cervone, G., Clemente-Harding, L., Alessandrini, S., and Delle Monache, L.: Short-term photovoltaic power forecasting using Artificial Neural Networks and an Analog Ensemble, Renewable energy, 108, 274–286, 2017.

Champagne, F. H., Friehe, C. A., LaRue, J. C., and Wynagaard, J. C.: Flux measurements, flux estimation techniques, and fine-scale turbulence measurements in the unstable surface layer over land, Journal of the Atmospheric Sciences, 34, 515–530, 1977.

Clifton, A., Kilcher, L., Lundquist, J., and Fleming, P.: Using machine learning to predict wind turbine power output, Environmental research letters, 8, 024 009, 2013.

Coen, J. L., Cameron, M., Michalakes, J., Patton, E. G., Riggan, P. J., and Yedinak, K. M.: WRF-Fire: coupled weather–wildland fire modeling with the weather research and forecasting model, Journal of Applied Meteorology and Climatology, 52, 16–38, 2013.

De Franceschi, M. and Zardi, D.: Evaluation of cut-off frequency and correction of filter-induced phase lag and attenuation in eddy covariance analysis of turbulence data, Boundary-Layer Meteorology, 108, 289–303, 2003.

Fernando, H. J., Mann, J., Palma, J. M., Lundquist, J. K., Barthelmie, R. J., Belo Pereira, M., Brown, W. O., Chow, F. K., Gerz, T., Hocut, C. M., et al.: The Perdigão: Peering into Microscale Details of Mountain Winds, Bulletin of the American Meteorological Society, 2018.

Frehlich, R.: Coherent Doppler lidar signal covariance including wind shear and wind turbulence, Applied Optics, 33, 6472–6481, 1994.

Frehlich, R., Meillier, Y., Jensen, M. L., Balsley, B., and Sharman, R.: Measurements of boundary layer profiles in an urban environment, Journal of Applied Meteorology and Climatology, 45, 821–837, 2006.

Friedman, J., Hastie, T., and Tibshirani, R.: The elements of statistical learning, vol. 1, Springer series in statistics New York, 2001.

Frisch, U.: Turbulence: the legacy of A.N. Kolmogorov, Cambridge University Press, 1995.

Garratt, J. R.: The atmospheric boundary layer, Earth-Science Reviews, 37, 89–134, 1994.

https://doi.org/10.5194/gmd-2020-16
Preprint. Discussion started: 21 April 2020

Gentine, P., Pritchard, M., Rasp, S., Reinaudi, G., and Yacalis, G.: Could machine learning break the convection parameterization deadlock?, Geophysical Research Letters, 45, 5742–5751, 2018.

Géron, A.: Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems, " O'Reilly Media, Inc.", 2017.

Gerz, T., Holzäpfel, F., Bryant, W., Köpp, F., Frech, M., Tafferner, A., and Winckelmans, G.: Research towards a wake-vortex advisory system for optimal aircraft spacing, Comptes Rendus Physique, 6, 501–523, 2005.

Giebel, G., Brownsword, R., Kariniotakis, G., Denhard, M., and Draxl, C.: The state-of-the-art in short-term prediction of wind power: A literature overview, ANEMOS. plus, 2011.

Hoerl, A. E. and Kennard, R. W.: Ridge regression: Biased estimation for nonorthogonal problems, Technometrics, 12, 55–67, 1970.

Hong, S.-Y. and Dudhia, J.: Next-generation numerical weather prediction: Bridging parameterization, explicit clouds, and large eddies, Bulletin of the American Meteorological Society, 93, ES6–ES9, 2012.

Huang, K., Fu, J. S., Hsu, N. C., Gao, Y., Dong, X., Tsay, S.-C., and Lam, Y. F.: Impact assessment of biomass burning on air quality in Southeast and East Asia during BASE-ASIA, Atmospheric Environment, 78, 291–302, 2013.

Kelley, N. D., Jonkman, B., and Scott, G.: Great Plains Turbulence Environment: Its Origins, Impact, and Simulation, Tech. rep., National Renewable Energy Laboratory (NREL), Golden, CO., https://www.nrel.gov/docs/fy07osti/40176.pdf, 2006.

Kolmogorov, A. N.: Dissipation of energy in locally isotropic turbulence, in: Dokl. Akad. Nauk SSSR, vol. 32, pp. 16–18, 1941.

Krishnamurthy, R., Calhoun, R., Billings, B., and Doyle, J.: Wind turbulence estimates in a valley by coherent Doppler lidar, Meteorological Applications, 18, 361–371, 2011.

Leahy, K., Hu, R. L., Konstantakopoulos, I. C., Spanos, C. J., and Agogino, A. M.: Diagnosing wind turbine faults using machine learning techniques applied to operational data, in: 2016 IEEE International Conference on Prognostics and Health Management (ICPHM), pp. 1–8, IEEE, 2016.

Leufen, L. H. and Schädler, G.: Calculating the turbulent fluxes in the atmospheric surface layer with neural networks, Geoscientific Model Development Discussions, 2018, 1–22, https://doi.org/10.5194/gmd-2018-263, https://www.geosci-model-dev-discuss.net/gmd-2018-263/, 2018.

Lundquist, J. K. and Bariteau, L.: Dissipation of Turbulence in the Wake of a Wind Turbine, Boundary-Layer Meteorology, 154, 229–241, https://doi.org/10.1007/s10546-014-9978-3, http://link.springer.com/10.1007/s10546-014-9978-3, 2015.

Mellor, G. L. and Yamada, T.: A hierarchy of turbulence closure models for planetary boundary layers, Journal of the Atmospheric Sciences, 31, 1791–1806, 1974.

Monin, A. S. and Obukhov, A. M.: Basic laws of turbulent mixing in the surface layer of the atmosphere, Contrib. Geophys. Inst. Acad. Sci. USSR, 151, 1954.

Muñoz-Esparza, D., Cañadillas, B., Neumann, T., and van Beeck, J.: Turbulent fluxes, stability and shear in the offshore environment: Mesoscale modelling and field observations at FINO1, Journal of Renewable and Sustainable Energy, 4, 063 136, https://doi.org/10.1063/1.4769201, 2012.

Muñoz-Esparza, D., Sharman, R. D., and Lundquist, J. K.: Turbulence dissipation rate in the atmospheric boundary layer: Observations and WRF mesoscale modeling during the XPIA field campaign, Monthly Weather Review, 146, 351–371, 2018.

Nakanish, M.: Improvement of the Mellor–Yamada turbulence closure model based on large-eddy simulation data, Boundary-layer meteorology, 99, 349–378, 2001.

https://doi.org/10.5194/gmd-2020-16
Preprint. Discussion started: 21 April 2020

Geoscientific Model Development Discussions

Nakanishi, M. and Niino, H.: An improved Mellor–Yamada level-3 model: Its numerical stability and application to a regional prediction of advection fog, Boundary-Layer Meteorology, 119, 397–407, 2006.

Oncley, S. P., Friehe, C. A., Larue, J. C., Businger, J. A., Itsweire, E. C., and Chang, S. S.: Surface-layer fluxes, profiles, and turbulence measurements over uniform terrain under near-neutral conditions, Journal of the Atmospheric Sciences, 53, 1029–1044, 1996.

Optis, M. and Perr-Sauer, J.: The importance of atmospheric turbulence and stability in machine-learning models of wind farm power production, Renewable and Sustainable Energy Reviews, 112, 27–41, 2019.

Paquin, J. E. and Pond, S.: The determination of the Kolmogoroff constants for velocity, temperature and humidity fluctuations from second- and third-order structure functions, Journal of Fluid Mechanics, 50, 257–269, 1971.

Sharma, N., Sharma, P., Irwin, D., and Shenoy, P.: Predicting solar generation from weather forecasts using machine learning, in: 2011 IEEE International Conference on Smart Grid Communications, pp. 528–533, IEEE, 2011.

Shaw, W. J. and LeMone, M. A.: Turbulence dissipation rate measured by 915 MHz wind profiling radars compared with in-situ tower and aircraft data, in: 12th Symposium on Meteorological Observations and Instrumentation, https://ams.confex.com/ams/pdfpapers/58647.pdf, 2003.

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Wang, W., and Powers, J. G.: A description of the advanced research WRF version 2, Tech. rep., National Center For Atmospheric Research, Boulder, CO, Mesoscale and Microscale Meteorology Div, 2005.

Smalikho, I. N.: On measurement of the dissipation rate of the turbulent energy with a cw Doppler lidar, ATMOSPHERIC AND OCEANIC OPTICS C/C OF OPTIKA ATMOSFERY I OKEANA, 8, 788–793, 1995.

Sreenivasan, K. R.: On the universality of the Kolmogorov constant, Physics of Fluids, 7, 2778–2784, 1995.

Thobois, L. P., Krishnamurthy, R., Loaec, S., Cariou, J. P., Dolfi-Bouteyre, A., and Valla, M.: Wind and EDR measurements with scanning Doppler LIDARs for preparing future weather dependent separation concepts, in: 7th AIAA Atmospheric and Space Environments Conference, p. 3317, 2015.

UCAR/NCAR, E. O. L.: NCAR/EOL Quality Controlled High-rate ISFS surface flux data, geographic coordinate, tilt corrected. Version 1.1, https://doi.org/10.26023/8x1n-tct4-p50x, 2019.

Wilczak, J. M., Oncley, S. P., and Stage, S. A.: Sonic anemometer tilt correction algorithms, Boundary-Layer Meteorology, 99, 127–150, 2001.

Wildmann, N., Bodini, N., Lundquist, J. K., Bariteau, L., and Wagner, J.: Estimation of turbulence parameters from scanning lidars and in-situ instrumentation in the Perdigão 2017 campaign, Atmospheric Measurement Techniques Discussion, in review, –, https://doi.org/10.5194/amt-2019-171, 2019.

Xingjian, S., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., and Woo, W.-C.: Convolutional LSTM network: A machine learning approach for precipitation nowcasting, in: Advances in Neural Information Processing Systems, pp. 802–810, 2015.

Yang, B., Qian, Y., Berg, L. K., Ma, P.-L., Wharton, S., Bulaevskaya, V., Yan, H., Hou, Z., and Shaw, W. J.: Sensitivity of turbine-height wind speeds to parameters in planetary boundary-layer and surface-layer schemes in the weather research and forecasting model, Boundary-Layer Meteorology, 162, 117–142, 2017.