Received: 23 April 2021 | Revised: 16 November 2021 | Accepted: 30 January 2022

DOI: 10.1002/we.2720

RESEARCH ARTICLE

WILEY

Retrospective de-trending of wind site turbulence using machine learning

Fraser Tough[1,2] | **Edward Hart**[3]

[1]Renewable Energy Systems (RES) Ltd., Kings Langley, UK

[2]School of Mathematics and Statistics, University of Glasgow, University Place, Glasgow, UK

[3]Wind Energy and Control Centre, Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow, UK

Correspondence
Edward Hart, Wind Energy and Control Centre, Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow, UK.
Email: edward.hart@strath.ac.uk

Abstract

This paper considers the removal of low-frequency trend contributions from turbulence intensity values at sites for which only 10-min statistics in wind speed are available. It is proposed the problem be reformulated as a direct regression task, solvable using machine learning techniques in conjunction with training data formed from measurements at sites for which underlying (non-averaged) wind data are available. Once trained, the machine learning models can de-trend sites for which only 10-min statistics have been retained. A range of machine learning techniques are tested, for cases of linear and filtered approaches to de-trending, using data from 14 sites. Results indicate this approach allows for excellent approximation of de-trended turbulence intensity distributions at unobserved sites, providing significant improvements over the existing recommended method. The best results were obtained using Neural Network, Random Forest and Boosted Tree models.

KEYWORDS
turbulence, de-trending, resource assessment, site conditions, machine learning, wind energy

1 | INTRODUCTION

As existing wind farm assets age, decisions regarding whether to decommission, life-extend and/or repower have important implications for project finances.[1–3] A crucial aspect of such decision making is an operators' ability to assess remaining fatigue life of turbine components. These assessments rely on load simulations which in turn require information regarding turbulence levels at the site in question. A turbulence intensity distribution, derived from data collected at the site, allows for representative wind fields to be generated and used as inputs to fatigue assessment simulations. A key issue when characterising turbulence in the atmospheric boundary layer (ABL) is the fact that fluctuations are present across almost all scales. This includes fatigue driving micro-scale contributions (turbulence), along with lower frequency sub-mesoscale and mesoscale contributions. Characterisation of the ABL, including the separation of turbulence from other components, has been extensively studied in the boundary-layer meteorology literature; for example, see Caughey,[4] Smedman,[5] Vickers and Mahrt,[6] and Basu et al.[7] These works highlight the complexity of the atmospheric flow with which wind turbines interact. For many wind farm sites, a decade or more's worth of wind data is available for the assessment of site turbulence; however, this is commonly in the form of 10-min means and standard deviations computed from the raw wind speed measurements. For reasons which will be outlined in Section 2.1, wind speed variance values computed from measurements containing non-turbulence components lead to inflated fatigue estimates[8] and, hence, overly conservative remaining-life

Abbreviations: CV, coefficient of variation; TI, turbulence intensity; NN, neural network; RF, random forest; BT, boosted trees; kNN, k nearest-neighbours; MARS, multivariate adaptive regression splines; RMSE, root mean squared error; EMD, earth-movers distance; MR, mean ratio; ODT, over-de-trending.

estimates. Where raw data are available, these shortcomings can be avoided by separating micro- and macro-scale wind speed components, before computing summary statistics for the former only. The lower frequency (macro-scale) components are commonly interpreted as a time-varying trend[8,9] in the wind speed signal. This interpretation is adopted in the current work. In this context, the problem becomes that of *de-trending* the wind speed signal prior to calculating 10-min mean and standard deviation values. Due to the non-stationarity of wind speed time-series, robust filtering methods (such as a wavelet transform or empirical mode decomposition) are preferable to Fourier-based filtering when seeking to separate turbulence and trend contributions.[6,7] However, at present, the most commonly utilised methods for trend identification, in the context of site assessment for load analysis, are linear de-trending (i.e., performing straight line fits) and Fourier-based filtering.[8–10] Both approaches feature in current IEC standards,[11,12] although others are permitted. While motivations for de-trending, outlined above, have thus far been based on decisions made towards the end of a project's lifetime, it is important to note that turbulence characterisation also influences design stage decision making. As awareness of the need for de-trending grows, it is increasingly common for it to be incorporated at the data collection stage. However, this is not always the case, and so the current problem remains relevant to the design stages of some new wind farms.

In cases where only moment data (i.e., means and standard deviations) have been retained, direct trend removal is no longer an option. Methods which allow retrospective de-trending using moment only data have therefore been developed. Model 2 from Larsen and Hansen[8] in particular demonstrated promising results. More recently, however, the viability of this method was brought into question,[9] resulting in a recommendation that it no longer be used. Comparisons of a Gaussian process approach[9] with Model 1 of Larsen and Hansen[8] indicated that the remaining viable methods provide only very conservative estimates of de-trended turbulence distributions. Therefore, improvements in retrospective de-trending capabilities have the potential to unlock significant value for the wind industry. This paper aims to facilitate the required improvements by proposing a reformulation of the de-trending problem such that it can be tackled using machine learning methods. A strong emphasis is placed on practicality, accessibility and easy uptake by operators, with all analysis therefore performed using open-source software packages. For these same reasons, the present study focuses on the commonly used linear and Fourier based (filtering) de-trending methods. However, the presented methodology is general and, as will be shown, can be applied for different choices of de-trending method.

Note, to distinguish between data types, raw wind speed measurements at their original capture frequency will be referred to as *high-frequency data* (irrespective of the actual capture frequency); 10-min statistical quantities will be referred to as *moment data*.

Section 2 presents a comprehensive introduction to the problem and related research carried out to date. Section 3 highlights the distribution of reductions in TI obtained when de-trending is undertaken on high-frequency data. Section 4 then proposes a reformulation of retrospective de-trending as a regression task. The employed data and data preparation procedures are outlined in Section 5, with implemented machine learning models then introduced in Section 6. Results are presented in Section 7 before a discussion and conclusions close the paper.

2 | BACKGROUND

Atmospheric wind is responsible for not only the generation of wind energy, but also fatigue loading on wind turbine components. Random variations in wind speed causes damage to blades, towers and drivetrains, reducing the lifetimes of these components and that of each turbine as a whole. These variations in wind speed are known as turbulence. Different classes of turbine are built to withstand different levels of turbulence—with higher class turbines designed to withstand more extreme conditions. These conditions must be accurately quantified prior to development of a wind farm so that developers can (a) determine the appropriate class of turbine to erect on site and (b) determine the expected lifetime of the chosen class for that site. As discussed, these same considerations are again important later in a projects lifetime, when decisions surrounding decommissioning and life extension are made. These design and expected lifetime considerations have significant cost implications for projects and, thus, represent an opportunity for improving cost efficiencies. Furthermore, turbulence related curtailment* can be necessary during wind farm operation, driven by the total turbulence being experienced. Total turbulence for a given turbine is an aggregate of the ambient turbulence plus wake added turbulence, caused by turbines sitting within wake sectors of upwind machines. Characterisation of ambient turbulence therefore also influences wind farm operational decisions.

Generally, to quantify site conditions, meteorological masts are erected at least 1 year before the development of a project and data is then collected throughout the project life. High-frequency measurements are collected (generally at between 1 and 50 Hz) and, for each consecutive 10-min period the mean (μ) and residual standard deviation from the mean (σ), are computed. In addition, relative variation in wind speed can be expressed using the Coefficient of Variation (CV), defined as $CV = \sigma/\mu$. Once each 10-min data point has been calculated, the high-frequency data

*Curtailment being when wind turbine power outputs across a wind farm are reduced below their maximum potential levels given the current wind conditions. Curtailment may be employed to improve load conditions for turbine components (as is the case for instances of very high turbulence), to support grid stability or to improve wind farm efficiency. Curtailment is achieved through control actions, these being adjustments to generator torque or blade pitch angle. A number of different curtailment strategies have been developed and investigated; for example, see Galinos et al,[13] and Lio et al.[14]

are discarded, and only the 10-min statistics are retained. This process is carried out due to the high cost of storing raw (high-frequency) data. The CV is a general statistical term for relative variation. In the wind industry, CV is referred to as Turbulence Intensity (TI).

2.1 | De-trending turbulence intensity - high-frequency data

The problem with using TI to generate wind fields in which load variations are estimated is the implicit assumption that wind speed is a stationary process (a time-series with constant mean and variance). However, wind speed time-series often exhibit behaviour of a non-stationary process, specifically that the conditional mean is not constant over time but is instead slowly varying. The presence of these slowly varying trends inflates load estimates because, under the assumption of stationarity in turbulence generating codes, the energy associated with increased variance levels (caused by the presence of low-frequency trends) is re-distributed across the whole turbulence spectrum.[8] As such, fatigue damage estimates from simulations are inflated relative to that which would result from operation in the true wind field from which measurements were obtained. Such trend contributions should therefore ideally be removed prior to computing TI.

Little research has been carried out to link fatigue damage with the most appropriate degree of smoothing during trend identification. Linear de-trending within each 10-min interval provides a conservative approach. Hansen and Larsen[15] assessed the effect of linear de-trending, utilising data from masts across northern Europe, predominantly Sweden, Denmark, the Netherlands and Norway. Sites were both offshore and coastal. The linear de-trending method reduced TI by between 3% and 15% across these sites. Trend contributions were generally higher for coastal compared to offshore sites, but trend was found to be fairly constant across the wind speed range. While simple to implement, linear de-trending removes only part of the slowly varying component of TI. Damaschke et al[16] investigated the use of methods which provided a more flexible fit to the data. Four different methods were compared: constant mean; a form of non-parametric regression essentially splitting the data up into ten 1-min intervals assuming a constant mean within each; a high-pass filtering method[12] and polynomial regression (although no information on the degree of polynomial is provided). TI was estimated at 14% when applying no de-trending, 8% using polynomial regression, 7% using non-parametric averaging and finally 6% using high-pass filtering. Little information on the data is provided, although it is inferred a large database was utilised as well as simulated data.

In Larsen and Hansen,[8] a more rigourous definition of trend was adopted, in which the signal trend is defined to consist of the sum of all Fourier modes present in the signal with period (equivalently wavelength, since this is a time-series signal) greater than the capture window over which statistics are calculated (10 min). This definition coincides with that of previous work into atmospheric boundary layer flows.[9] In a fatigue context, the definition makes intuitive sense, since fatigue is driven by cycles of loading which in turn are driven by cycles in wind speed; across a given capture window, the stochastic, fatigue contributing component, is defined to be all frequency components that complete at least one full cycle, with the remainder defined as the slowly varying and non-cyclical (within the measurement window at least) trend. This definition of trend was also used in Hart et al[9] and is again adopted in the current paper. As in the previous work, trend portions of measured wind speed time-series are identified here via a third order Butterworth low-pass filter with cutoff frequency set to 1/600 Hz. Straight-line fits are also used to identify linear trends in wind speed data since, as has been outlined, this remains common practise. Results for both linear and filtered de-trending approaches will be presented, in some sense representing the lower and upper levels of variance which may be removed from within 10-min capture windows. Visual examples of the residuals resulting from different approaches to trend fitting are shown in Figure 1; also see Hart et al.[9] It is emphasised that there is no single definition of *trend* which applies absolutely or universally to this problem; the chosen definition was adopted due to its precedence based on previous work and its intuitive nature, as described above. De-trending based on this definition, in combination with linear de-trending, was deemed appropriate and sufficient for the development of machine learning based moment de-trending techniques. In the future, this definition of trend may be refined or improved, but in order to do so, further work will be necessary. These ideas are revisited in Section 8.

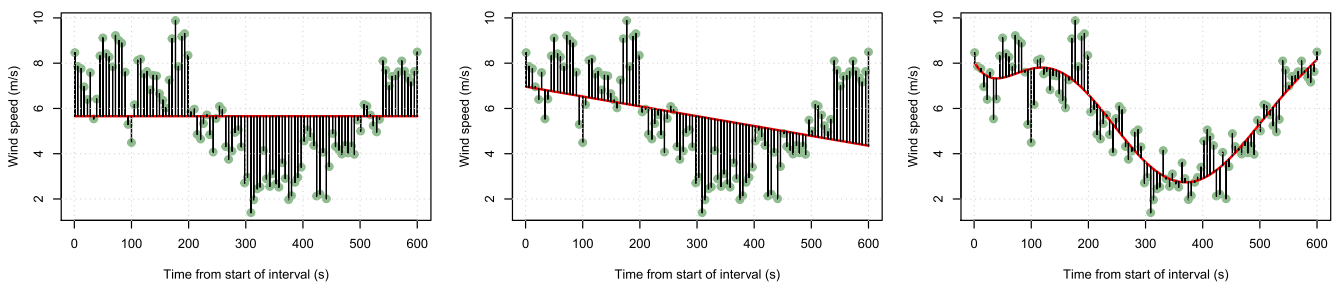

FIGURE 1 Wind speed residuals versus time for different possible trend fits: (left) no trend fit—0th degree polynomial; (middle) straight line—1st degree polynomial; (right) 3rd degree polynomial

2.2 | De-trending turbulence intensity- moment data

While manufacturers and wind farm developers have cumulatively collected thousands of years' worth of 10-min statistics, the high-frequency data required to compute de-trended TI (TI^{dt}) has generally been discarded due to high storage costs. Furthermore, the value in storing high-frequency data was historically unknown. Loggers can be adapted to collect high-frequency data or even perform de-trending in situ; however, it will take considerable time for significant quantities of de-trended TI data to be collected across existing sites. In addition, it is cost effective to use information that is currently available. For these reasons, a number of post-processing methodologies have been developed. These aim to remove trend contributions from statistical data when only 10-min first and second order moments (μ and σ) are available. Hansen and Larsen[15] developed an empirical model which depended on the current standard deviation, current mean and lead mean. However, Damaschke et al[16] compared the performance of this model with true linear de-trending, concluding that the empirical model had little effect (estimating a reduction of 0.3%, compared to the true reduction of 3.1%). In Larsen and Hansen,[8] two separate moment de-trending models (Models 1 and 2) are presented, both of which fit continuous spline-like polynomials to 10-min statistics in order to estimate trends. Linear trend contributions are then removed via analytical equations. A model comparison in the paper utilises 10 high-frequency datasets across Denmark, Norway and the USA, concluding that Model 2 is the superior method and recommending its use for de-trending of wind site data. Hart et al[9] reexamined the Larsen and Hansen methods, while also developing a Gaussian process (GP) regression approach in which smoothing levels were set using trend components in real data. It was shown that the iterative solving technique suggested for Model 2 generally fails to find the globally optimal solution to the fitting equation system; furthermore, when true best-fits were found, the method was shown to over-fit, consistently removing a much larger proportion of TI than is permissible. It was therefore recommended that Model 2 no longer be used as a de-trending technique. Comparisons between GP and Model 1 results showed that these two methods produce very similar outputs. In both cases, only conservative estimates of de-trended TI were obtained, with one possible implication being that improved results may require additional information (this point is revisited below). Resulting similarities in performance were interpreted as a validation of smoothing levels present in Model 1 and, owing to the additional theoretical complexities of the GP approach, Model 1 was recommended as the best de-trending method available at the time. Since currently viable methods are known to provide only very conservative estimates of de-trended TI, there is a clear requirement for improved moment de-trending capabilities.

3 | QUANTIFYING TI REDUCTIONS FROM HIGH-FREQUENCY DE-TRENDING

In addition to the main focus of moment de-trending, it was also of interest to quantify effects of linear and filtered de-trending on high-frequency datasets, as occurs if de-trending is undertaken in parallel to data collection at individual sites. Using all datasets (see Section 5) in the current study, 95% prediction intervals were determined for the percentage reduction in mean TI from both de-trending approaches on raw (high-frequency) data. For linear de-trending, the resulting interval is

$$(2.1\%, 14.8\%), \tag{1}$$

and in the filtered case,

$$(4.6\%, 22.2\%), \tag{2}$$

with central estimates of 7.0% and 11.7%, respectively. The 95% prediction intervals state that, given a new unobserved dataset, there is a 95% probability that the percentage reduction in TI will lie between 2.1% and 14.8% for linear de-trending and 4.6% and 22.2% for filtered de-trending. Central estimates provide the most likely value to be observed in each case. Full details of the process by which intervals were determined are given in Appendix A.

4 | A REFORMULATION OF THE DE-TRENDING PROBLEM

There is a need for improved de-trending methodologies in the case of moment data and, in particular, a need for methods which provide more accurate, less conservative estimates of de-trended TI. However, as highlighted in the previous section, and discussed in Hart et al,[9] existing methods may already be performing as well as is possible given the information available to them. More explicitly, existing approaches can all be described as *time-series trend identification models*, in which the trend portion of raw data is approximated via curve fitting on concurrent moment values, with trend variance then subtracted from measured variance to obtain de-trended TI. It is important to consider the fact that in forming statistical moments, a large number (≥600) of high-frequency measurements are summarised using just two representative 10-min values, μ and σ.

Furthermore, fitting models to concurrent values means only a few leads and lags contribute to the approximation of wind speed trend within any individual 10-min window, as such, this is not a formulation in which results can be improved by simply using more data. Hence, the information loss associated with converting measurements to first and second order moments cannot, in this case, be counteracted through the provision of larger datasets. It therefore seems sensible to expect that a performance ceiling will exist for the current problem formulation, regardless of the specifics of the implementation.

In order to avoid such issues and move to a scenario in which training data quantities do link directly to model accuracy, it is proposed that the retrospective turbulence de-trending problem be reformulated as a *direct regression problem* between TI and TI^{dt}. In this form, one can use available high-frequency data across a range of sites to calculate first and second order moments as well as TI and TI^{dt}. Machine learning models can then be used to relate TI^{dt} to other variables (including TI) that are generally available in moment datasets. The model(s), once developed, can then be utilised at sites where only moment data is available, allowing TI^{dt} distributions to be estimated.

The proposed TI de-trending regression models are of the following form,

$$TI_t^{dt} = f(TI_{t-5}, ..., TI_{t+5}, \mu_{t-5}, ..., \mu_{t+5}, \sigma_{t-5}, ..., \sigma_{t+5}, \text{Terrain}), \tag{3}$$

where indices indicate the current time point (t) and lead/lag time points ($t \pm i$). Terrain is a categorical variable which indicates the terrain type present at a given site. f denotes the unknown function, to be identified via machine learning, which relates the inputs to the de-trended output values. TI^{dt} values on which to train can be generated using any choice of de-trending method. De-trended TI is therefore modelled as a function of concurrent values of un-detrended TI, mean wind speed and un-detrended wind speed standard deviation, five leads and lags of these same variables and an indicator of terrain type. The variables μ and σ contain more information than TI alone, since the same value of TI can be generated by many combinations of μ and σ. While TI itself is determined by μ and σ, it was found to be beneficial to include TI itself since the complexity of resulting models reduces. This is because the relationship between TI^{dt} and TI is approximately linear. It was found that wind speed measurement frequency was not an important variable when predicting TI reductions; frequency was therefore not included as an input to regression models. All input variables were transformed to Z scores[†] prior to modelling.

While the described linear and filtered de-trending approaches are studied here, the proposed regression approach is general and not restricted to these. The presented methodology can be used to estimate de-trended TI for any chosen de-trending method. All that is required are resulting de-trended TI values on which to train the model, obtained by applying the chosen method to raw data, as well as the associated un-detrended statistics TI, μ and σ (along with other possible inputs to the regression problem one might wish to apply).

5 | DATASETS AND DATA PREPARATION

To develop data driven models, high-frequency wind data were required. Data from 14 sites captured in the Technical University of Denmark's (DTU) Database on Wind Characteristics[‡] was used for the study. For each site, individual data files were combined to form continuous time-series of wind speed measurements at each available height. De-trended time-series were generated by (a) linear de-trending and (b) low-pass filtering. Low-pass filtering was carried out using a cut-off frequency corresponding to 10-min capture windows (see Section 2.1) and then subtracting the identified trend from measured values.[§] Three sets of TI values were then generated (all for 10-min capture windows): *un-detrended* TI using raw and unaltered data, *linearly de-trended* TI by taking residuals from a straight line fit within each window and, finally, *filter de-trended* TI using the residuals obtained by subtracting the filter identified trend from measured data. Note, when de-trending TI, it is only standard deviation values, σ, which are adjusted. The 10-min mean values, μ, remain the same regardless of whether a trend is present or not. Since lead and lag values are required inputs, data concurrency checks were carried out to ensure regression datapoints were only generated where all required lead and lag values were available. In order to ensure equal weightings across sites, and hence avoid data quantity biases, a subset of 8,669 regression datapoints (the number of rows of the smallest dataset) was taken at random from the full set generated for each site. Finally, since TI is the ratio of σ to μ, small denominators relative to numerators can result in extremely large TI values. It was found that these extreme observations have a strong influence on model form and a detrimental effect on model accuracy. It was therefore necessary for such points—defined as the top 1% across all input variables and all datasets—to be removed from the data prior to the modelling stage. A summary of the 14 datasets is provided in Table 1.

[†]The standard score, or Z score, of each value taken by a given variable $x \in X$ (where X is the set of all observed values of this variable) is $Z = (x - \mu_X)/\sigma_X$ where μ_X and σ_X are, respectively, the mean and standard deviation of values in X. Therefore, the Z score describes the number of standard deviations by which an individual measurement is above or below the mean of measured values for that variable. Converting to Z scores therefore centres and scales (also non-dimensionalising) the values on which regression is to be performed. Centring and scaling in this way is known to reduce numerical instabilities, improve interpretability and avoid issues associated with multicollinearity when interaction terms are present.[17,18]

[‡]www.winddata.com

[§]Note one could instead simply apply a high-pass filter with the same cut-off frequency and avoid having the additional operation of subtracting the identified trend.

TABLE 1 Site data summary information

Site name	Dataset size (Num. 10-min stats)	Meas. heights (m)	Meas. freq. (Hz)	Terrain type
Ainswort	8,538	20, 30, 40	5	Complex
Cabauw	8,058	40, 80, 140, 200	2	Pastoral
Calwind	8,181	20, 30, 40	5	Complex
Flowind	7,441	20, 30, 40	5	Complex
Gedsrev	8,518	10, 30, 45	5	Offshore
Gorgonio	5,633	20, 30, 40	5	Complex
Hanford	6,445	20, 30, 40	5	Pastoral
Hurghada	7,811	29, 30	8	Pastoral
Lyse	8,069	11, 24, 32, 40	1.0069	Complex
Middelgrunden	8,502	10, 30, 45	5	Offshore
Orkney	8,138	60, 64	1	Coastal
Skipheia	8,004	≤100	0.8533	Coastal
Sletringen	7,968	≤45	0.8533	Coastal
Toboel	8,402	15, 30, 49	8	Pastoral
Total	**109,708**			

6 | MODELLING

Modelling aspects of the proposed de-trending reformulation will now be considered. The current section opens with a visual study of the relationship between TI^{dt} and two of its input covariates (μ and σ). This serves as a starting point for understanding the appropriate magnitude of de-trending for different combinations of the chosen variables and, importantly, provides insight into the functional form of relationships between input variables and de-trended TI. After this initial analysis follows an outline of the applied machine learning techniques, performance metrics and the training cross-validation procedure.

6.1 | Visualising relationships

Before training machine learning models across large numbers of input variables, it is valuable to first study relationships in the data for a small number of key variables. This allows one to gain insight regarding characteristics of the problem, and the final form models are likely to take. This is the goal of the current section.

Under the assumption of normally distributed errors, residual standard deviation from the *conditional mean* must result in a value equal to or less than residual standard deviation from the *mean*. Thus, assuming normality of errors, de-trended TI will always be less than TI. It is therefore reasonable to estimate and apply a simple scaling factor such that de-trended TI takes the form of un-detrended TI multiplied by a constant. This can be realised via a linear model consisting of a constant only, $y = \beta$. Estimating the parameter β using all available data results in estimates of $\beta = 0.93$ and $\beta = 0.88$ for linear and filtered de-trending, respectively. However, a single scale parameter results in an average proportional reduction. This will not be accurate for each and every point. Consider that, when wind speed magnitude is relatively constant over an interval, de-trended TI will be approximately equal to TI; thus, β will equal 1. However, when the wind speed is ramping up or down, the scaling factor will tend to decrease. If the model form is specified to allow for more flexibility, β can be allowed to vary as a function of other input variables in a non-linear way. These variations in β then allow the associated changes in de-trending magnitude to be captured and visualised. Varying coefficient models were developed for this purpose, using simple linear models in which β was allowed to vary as a function of discretised μ and σ values. The input variables were discretised in way which minimised overfitting.[¶] Figure 2 shows β as a function of μ and σ. Monotonically increasing behaviour can be observed.

When observing β while allowing μ and σ to interact with one another, the monotonically increasing form is not constant. The bivariate surfaces within Figure 3 were quantified by fitting a continuously varying coefficient model. Initial model fits suggested a monotonically increasing function in the μ dimension, as shown in Figure 3, with troughs in sparsely populated spaces. However, the σ dimension did not follow a strictly

[¶]Increasing the number of factor levels from the number shown here led to faceting of data in sparsely populated areas.

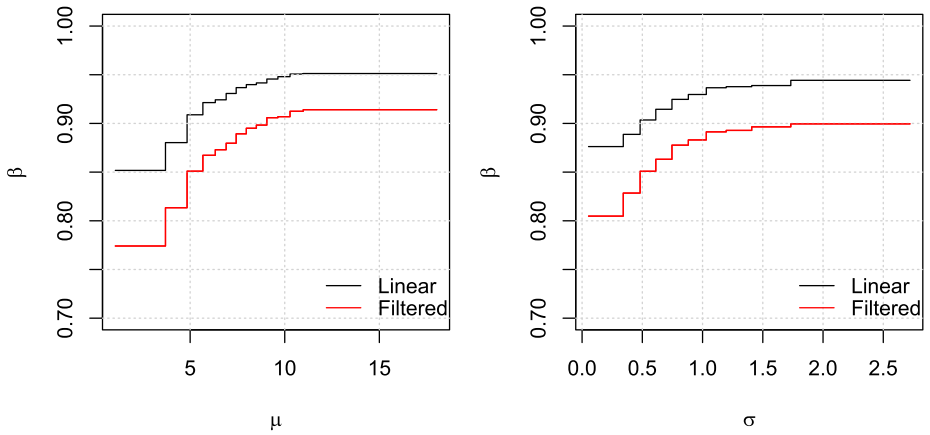

FIGURE 2 Scaling factor (β) as a function of mean (μ) and standard deviation (σ) of wind speed

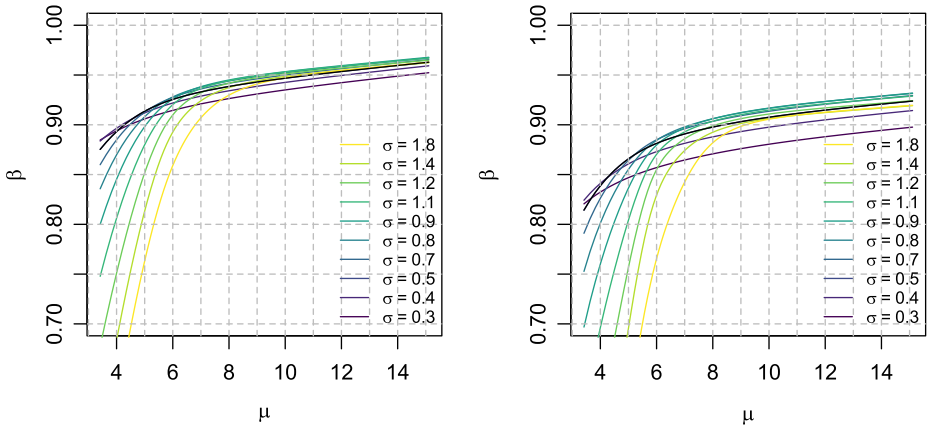

FIGURE 3 Scaling factor (β) joint variations with mean (μ) and standard deviation (σ) of wind speed for (left) linear de-trending (right) filtered de-trending

monotonically increasing behaviour. The bivariate varying coefficient model's form was therefore constrained such that β could only increase with increasing μ, removing the effect of sparsely populated regions on the form. The resulting surfaces indicate that when σ is low, reduction in TI is only weakly dependent on μ; that is, the reduction is near constant. However, as σ increases, TI reductions become increasingly strongly dependent on μ, especially at low mean wind speed values. This indicates that the largest reductions in TI from de-trending occur when σ is high and μ is low.

The relationships between β, μ and σ as observed in Figure 3 are not static when conditioned on further variables. The full hypersurface is therefore complex, varying in a non-linear way and not easily visualised in 2D or 3D. It is this complex, high-dimensional hypersurface which must be quantified to allow for accurate de-trending at sites where only 10-min data are available.

6.2 | Machine learning models

Machine learning models are used frequently within renewable energy[19] across a diverse range of applications, from power forecasting[20] to fatigue load estimation.[21] Alongside increased usage of such models has come associated improvements in accessibility, with underlying theory and computer implementations for most methods now readily available. For detailed considerations of the methods used here, excellent reference materials are available.[22,23]

TABLE 2 Regression models and associated R packages

Model	R package
Linear Model	stats, glmnet
Neural Network	nnet
Random Forest	ranger
Boosted Tree	xgboost, gbm
Multivariate Adaptive Regression Splines	earth, bagEarth
k-Nearest Neighbour	kknn

A wide range of machine learning models were assessed within the study to quantify the ability of each to identify the unknown functional relationship of Equation 3. These models have varying complexities and characteristics, as follows:

- **Linear Models** are models in which the practitioner must have an understanding of how variables relate so they can define the functional form of the relationship. This is carried out via specification of linear relations, polynomials and/or interaction variables. For standard linear models, there are no hyperparameters to optimise. However, regularised linear models were also assessed in the study. Regularised linear models protect from overfitting by either shrinking (ridge) coefficients or setting coefficients to 0 (lasso). Elastic net combines both ridge and lasso penalties. For elastic net regularised linear models, two hyperparameters are required, lambda and alpha. Lambda is the magnitude of regularisation, and alpha is the proportion of lasso relative to ridge regularisation.
- **Neural Networks** are learning algorithms which can identify functional relationships between inputs and outputs using networks of interconnected artificial neurons, mimicking aspects of biological brains. The complexity of neural networks depend on their architecture—the number of hidden layers and hidden units within layers. Feed forward neural networks were trained while optimising the number of hidden units, weight decay, the number of training iterations (epochs) and activation function (between the original predictors and the hidden unit layer).
- **Random Forests** are algorithmic models where large numbers of independent decision trees are fitted to the data. Predictions from individual trees are aggregated to form a final prediction. While training these models, three hyperparameters were optimised: the number of variables available for splitting at each tree node (mtry), the minimum number of data points in a node that are required for the node to be split further (min *n*) and the number of trees (trees).
- **Boosted Trees** are also algorithmic, where many decision trees are sequentially fitted to the data. The process of boosting is where a decision tree is fitted to the data, with another tree then fitted to the residuals, and so on. This process is carried out a large number of times. This algorithm was trained while optimising a wide range of hyperparameters: the maximum depth of the tree (tree depth), the rate that the algorithm adapts as it iterates (learn rate), the number required for the reduction in the loss function to split further (loss reduction), the minimum number of data points in a node that is required for the node to be split further (min *n*), the number of data points that are exposed to the fitting routine (sample size) and the number of trees contained in the ensemble (trees).
- **Multivariate Adaptive Regression Splines** are linear splines with matching endpoints that can capture non-linearities, splitting the data space into segments in which each linear segment is used to make predictions for inputs in that region. Hyperparameters that were optimised for this model were the number of features that were retained in the final model and the highest possible interaction degree.
- **K Nearest-Neighbours** is an algorithmic approach which estimates the response as a function of the inputs. It does this by computing a weighted average of the points surrounding an input location. While training this model, the number of neighbours, the weighting function and the Minkowski distance power were optimised.

Model types and their associated R packages are summarised in Table 2. Given the functional form of relationships observed between the investigated coefficient for TI and other covariates considered in Section 6.1, it is reasonable to assume that linear models with flexibility to allow for different β values as a function of inputs would provide improved fits over constant coefficient linear models. Therefore, three forms of linear model were assessed: one linear model served as a baseline (f1), including only TI as a predictor; a second linear model (f2) contained only linear main effects[#] and a third linear model (f3) was constructed which attempted to capture the full surface presented within Section 6.1, as well as other interactions between neighbouring statistics. A number of candidate model forms were assessed to arrive at the final form of f3. This model

[#]That is, effects of one variable are assumed independent of the other variables.
[‖]Where effects of one variable are dependent on the value of another.

included combinations of polynomials and varying degrees of interactions[‖] between them. The final form of f3 was chosen through trial and error as well as utilisation of elastic net regularisation.

Hyperparameter optimisation for all models was carried out within the cross-validation framework outlined in the following section. Hyperparameter optimisation was carried out via grid-search, with the same grid used for both linear and filtered de-trending models.

6.3 | Evaluating performance

Data were split into two datasets, training data and test data. The training dataset was used to determine optimal hyperparameters for each model. The test dataset was used to validate performance of the models after training. To split the data, data from each site was ordered in time from the least recent observations to most recent. The training dataset was defined as the least recent 70% of the data, and the test dataset was defined as the most recent 30%. Model development was carried out on the training dataset via cross-validation until each final model and its optimal hyperparameters were found. Final assessment of each model's performance was then conducted on observations from the most recent 30% of the data, which was not used for training. A cross-validation approach was required that would allow one to estimate representative out-of-sample error, as would be observed in practice. A leave-one-site-out (14 sites in total) cross validation procedure was therefore adopted. For example, one would train on sites 1–13 and predict for site 14, and then one would train on sites 1–12 and 14 and predict for site 13 and so on. Each training dataset represented 65% of the overall study database ($13/14 \times 70\%$), while each test dataset represented approximately 2% ($1/14 \times 30\%$). Model hyperparameters were optimised such that errors (see Section 6.4) were minimised across the full range of predicted datasets within the cross-validation procedure.

6.4 | Performance metrics

For each model in Table 2, hyperparameters which minimised Root Mean Squared Error (RMSE) between predicted and observed TI^{dt} were deemed optimal. Once hyperparameters were selected, the Earth-Movers Distance (EMD) and Mean-Ratio (MR) between predicted and observed TI^{dt} for each test set were also computed to provide a more detailed assessment of performance. For an outline of the EMD metric, see Hart et al.[9] Note that previous work used EMD to assess the 'quantity of de-trending' which had taken place; hence, maximal values in this metric were indicative of good performance. In the current case, EMD is instead used as a true error metric between predicted and observed TI^{dt} distributions, with the lowest EMD values therefore indicative of best performance here. The MR metric was also used, as defined in Larsen and Hansen,[8] where good performance is indicated by values being close to 1. Finally, the over-de-trending (ODT) measure of Hart et al[9] was applied to ensure resulting models do not reduce TI excessively.

7 | RESULTS AND COMPARISONS

Results for out-of-sample performance are presented in Tables 3 and 4 for linear and filtered de-trending cases, respectively. Results from Larsen and Hansen's M1 (linear de-trending) method are also included in Table 3. The chosen metrics allow for a detailed comparison of newly proposed machine learning models with the current state-of-the-art, Larsen and Hansen's Model 1 (M1), as per the literature.

TABLE 3 Out-of-sample performance across trained models (**linear** de-trending case)

Model	RMSE	WD	MR	RMSE rank
Baseline (f1)	0.0127	0.0033	1.0035	7
Linear (f2)	0.0110	0.0019	1.0026	5
Linear (f3)	0.0101	0.0015	1.0001	4
Neural network	0.0097	0.0013	1.0012	1
Random forest	0.0099	0.0014	0.9971	2
kNN	0.0153	0.0053	1.0125	8
MARS	0.0114	0.0023	1.0022	6
Boosted tree	0.0100	0.0014	0.9988	3
M1 (L&H)	0.0177	0.0345	1.0205	9

TABLE 4 Out-of-sample performance across trained models (**filtered** de-trending case)

Model	RMSE	WD	MR	RMSE rank
Baseline (f1)	0.0151	0.0025	0.9984	8
Linear (f2)	0.0124	0.0015	0.9995	5
Linear (f3)	0.0116	0.0012	1.0007	4
Neural network	0.0106	0.0011	1.0021	1
Random forest	0.0107	0.0015	0.9969	2
kNN	0.0131	0.0032	1.0134	6
MARS	0.0135	0.0030	0.9971	7
Boosted tree	0.0115	0.0019	1.0011	3

TABLE 5 Final model hyperparameters

Model	Hyperparameter	Value
Neural network	Hidden units	7
	Weight decay	1e−10
	Epochs	902
	Activation function	Linear
Random forest	Trees	875
	Mtry	15
	Min n	12
Boosted tree	Trees	883
	Min n	24
	Tree depth	2
	Learn rate	0.019
	Loss reduction	1.198
	Sample size	0.945

All three linear models were fitted with and without regularisation—regularisation reduced RMSE in all three cases, albeit negligibly for f1. Therefore, all linear models presented are regularised, with the effect of regularisation on out-of-sample metrics found to be most prominent for the complex linear model, f3, and least for the baseline model, f1, when compared to standard (un-regularised) linear models. Recall the baseline linear model (f1) only included TI as a predictor. The linear model which included only main effects (f2) can be seen to result in better out-of-sample metrics than the baseline model (f1), indicating that inclusion of additional variables over and above TI are useful. Performance for f2 can be seen to be inferior to that of f3, indicating the f2 model form was not flexible enough to quantify all complexity between inputs and the response. The third linear model, f3, included many more terms: main effects, polynomials and interactions. However, the minimum RMSE achieved was still higher than that of other tested models, namely Random Forest (RF), Boosted Tree (BT) and the Neural Network (NN). Therefore, the most advanced linear model form was still not flexible enough to characterise the full hypersurface. It is possible that, by defining higher dimensional relationships within the linear model framework, more accurate fits could be produced. However, at present, results indicate that de-trending functional forms between inputs and output are more complex than can easily be accounted for using a linear modelling approach.

An attempt was made to fit predictive Multivariate Adaptive Regression Splines (MARS). The function shown in Figure 3 can be approximated as a number of continuous linear sections which interact with each other in a nonlinear way. MARS models are simple, but effective when approximating surfaces that represent shapes with 'hinges'. However, MARS models are unstable predictors, and it is recommended bagging is used to exploit instabilities, much like how bagged regression trees result in a random forest.[**] It was thought that MARS might provide a reasonable fit in this case, as long as interaction terms were included. However, final metrics show that this model did not perform as well as RF, BT or the NN.

The k Nearest-Neighbours (kNN) algorithm did not perform well, most likely due to over-fitting. The specific kNN algorithm used does not have any regularisation parameters or constraints and simply computes a smooth average hypersurface based on the data alone. While a powerful method in some contexts, results indicate that the form of kNN assessed was ill suited to the problem.

**For example, see https://topepo.github.io/caret/miscellaneous-model-functions.html#bagMARS

RF and BT both performed well and can be recommended as candidate methods for moment de-trending. Note that two boosted trees were applied, these being gbm and xgboost. xgboost slightly outperformed gbm over the current datasets. The main difference between the two models is that xgboost is more heavily regularised than gbm and, thus, is slightly better at reducing over-fitting. Optimal hyperparameters for these models can be found in Table 5.

All other RMSEs and WDs exceed those obtained when using the NN, which was found to be the best performing model. This model was trained while optimising the number of hidden units, weight decay, the number of training iterations and activation function (between the original predictors and the hidden unit layer). Optimal hyperparameters can be found in Table 5. Note the same optimal RF, BT and NN hyperparameters were obtained for linear and filtered de-trending models. This is likely due to a combination of the standardisation of data values prior to modelling, see Section 4, and the use of grid-search with identical grids, see Section 6.2. It is expected that the use of finer grids would eventually result in different optimal hyperparameters for linear and filtered de-trending models.

A comparison of machine learning linear de-trending results with those of M1 (Table 3) indicates significant performance improvements have been realised across all metrics, with error reductions of 13–45% and 84–96% for RMSE and WD metrics, respectively. Across all sites, maximum percentage error reductions are for the NN model. RF and BT come in second and third positions, respectively.

Cumulative distribution functions of TI^{dt} and estimated TI^{dt} from the best performing model (NN) are shown in Figures 4 and 6 for linear and filtered de-trending cases, respectively, at four sites. M1 (linear de-trending) results at these same sites are shown in Figure 5 for comparison. Visually, it is clear that the presented machine learning approach to moment de-trending results in excellent approximations of the de-trended TI distributions obtained when high-frequency data are available. In particular, improvements over M1 are clear. Furthermore, these graphical results showcase the flexibility of the outlined approach, since excellent fits are also obtained in the filtered case, simply by adjusting the data on which models are trained.

ODT metric values were also assessed to ensure no excessive levels of over-de-trending are present across the test sites. ODT levels for the linear and filtered de-trending NN models across all 14 sites are shown in Figure 7. It can be seen that ODT levels, as defined in Hart et al,[9] remain

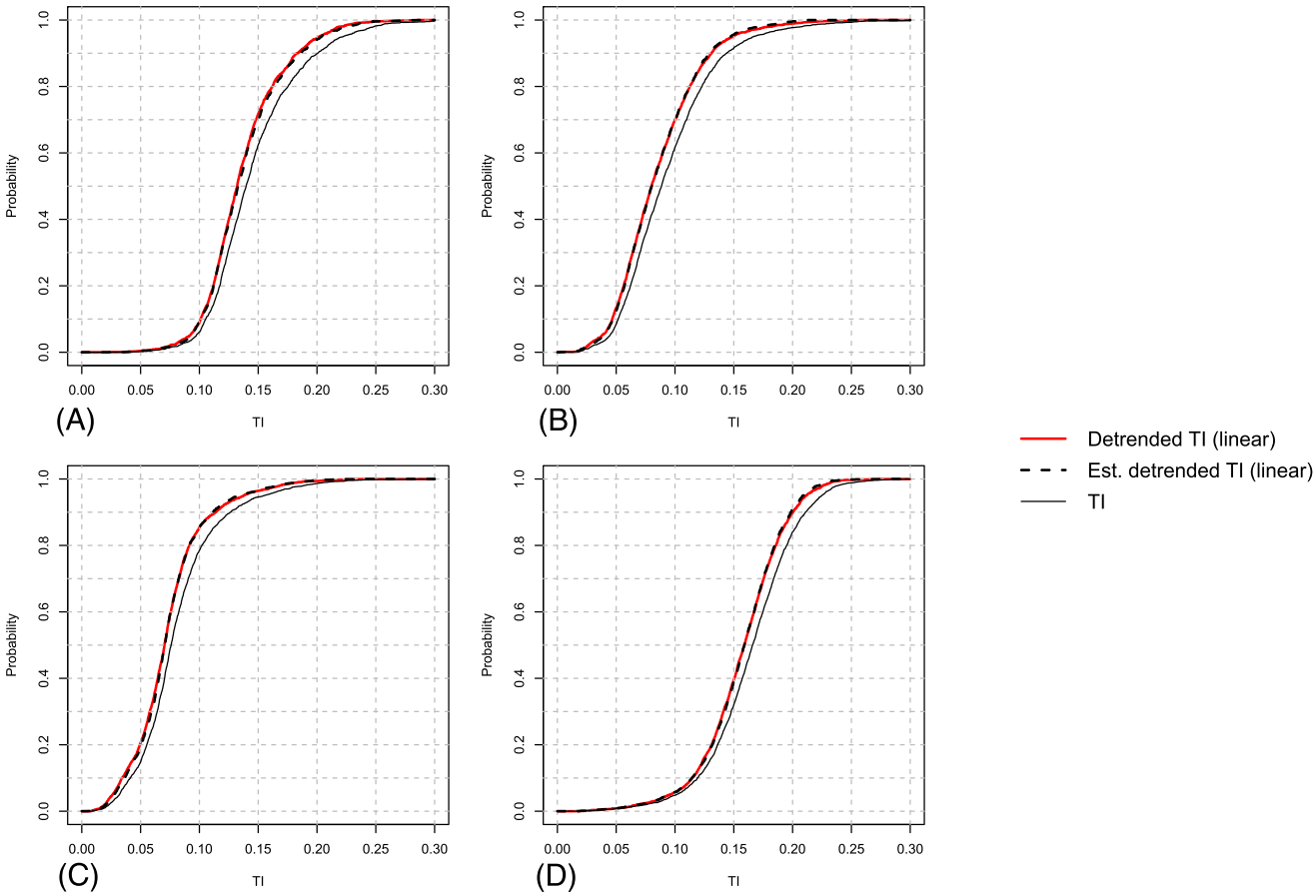

FIGURE 4 Cumulative distributions of TI in the linear de-trending case. Estimated TIs are those resulting from the best performing, NN, model. Results from four sites are shown: (A) Hanford, (B) Sletringen, (C) Orkney and (D) Toboel

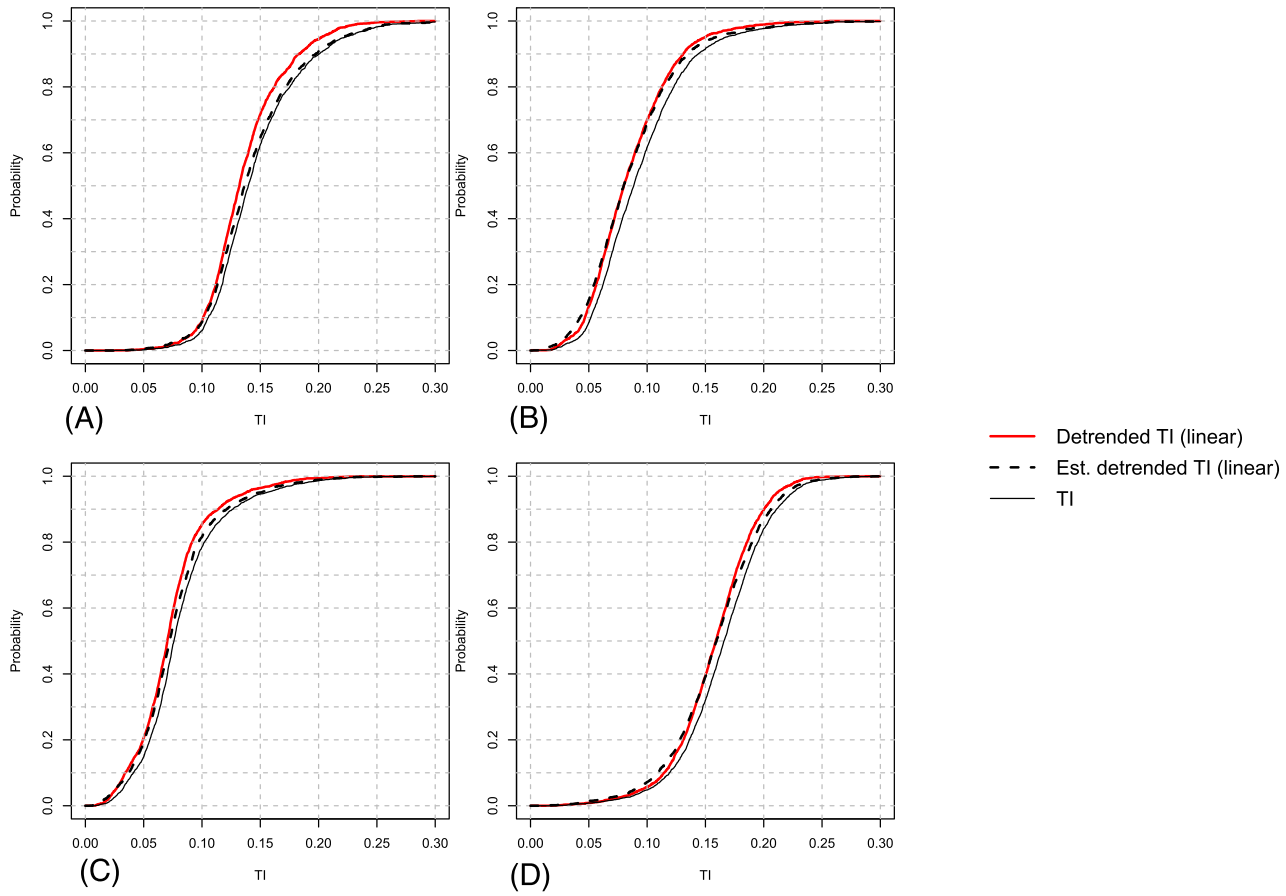

FIGURE 5 Cumulative distributions of TI in the linear de-trending case. Estimated TIs are those resulting from M1. Results from four sites are shown: (A) Hanford, (B) Sletringen, (C) Orkney and (D) Toboel

small in all cases, only increasing to just under 3.5% at a single site (Hurghada) and in most cases staying below 2%. When developing implementations of machine learning driven moment de-trending, ODT levels should remain a consideration during model development.

8 | DISCUSSION

Results of this work have demonstrated that machine learning based wind site moment de-trending can provide excellent approximations of the true de-trended turbulence distributions. An additional benefit of this approach is the flexibility to model varying degrees of de-trending, with good performance shown for both linear and filtered implementations. The best performing methods were found to be Neural Network, Random Forest and Boosted Tree models. It is important to note that generating accurate models is non-trivial, with careful consideration required regarding the preparation of training datasets and identification of an appropriate cross-validation procedure. With respect to the former, recall that extreme observations had to be removed prior to training in order to avoid reduced model accuracies. Regarding the latter, a representative cross-validation procedure is crucial for ensuring out-of-sample metrics are accurate and that chosen hyperparameters result in a generalised hypersurface, avoiding over-fitting. It is for this reason that full training datasets were 'left out' during cross-validation, as opposed to training and predicting on datasets containing disjoint subsets from the same site.

Regarding future work, it would be valuable to consider in more detail the utility of onshore and coastal data for de-trending near- and far-offshore sites. Such a study would require a more expansive dataset. Perhaps more importantly, now, it has been demonstrated that moment de-trending can be achieved for any method of de-trending; it is apposite to revisit the question of what may constitute appropriate/optimal de-trending, that is, that which leads to the most accurate fatigue life estimates for any given turbine or wind farm. Linear de-trending is known to provide conservative estimates, while the *Fourier modes* definition of trend applied in previous work,[8,9] and adopted again in the current paper, appears sensible and intuitive but remains a simplification. As discussed in Sections 1 and 2.1, the need for de-trending relates to simulation modelling assumptions of stationarity and the associated need to separate turbulence from sub-mesoscale and mesoscale components in site

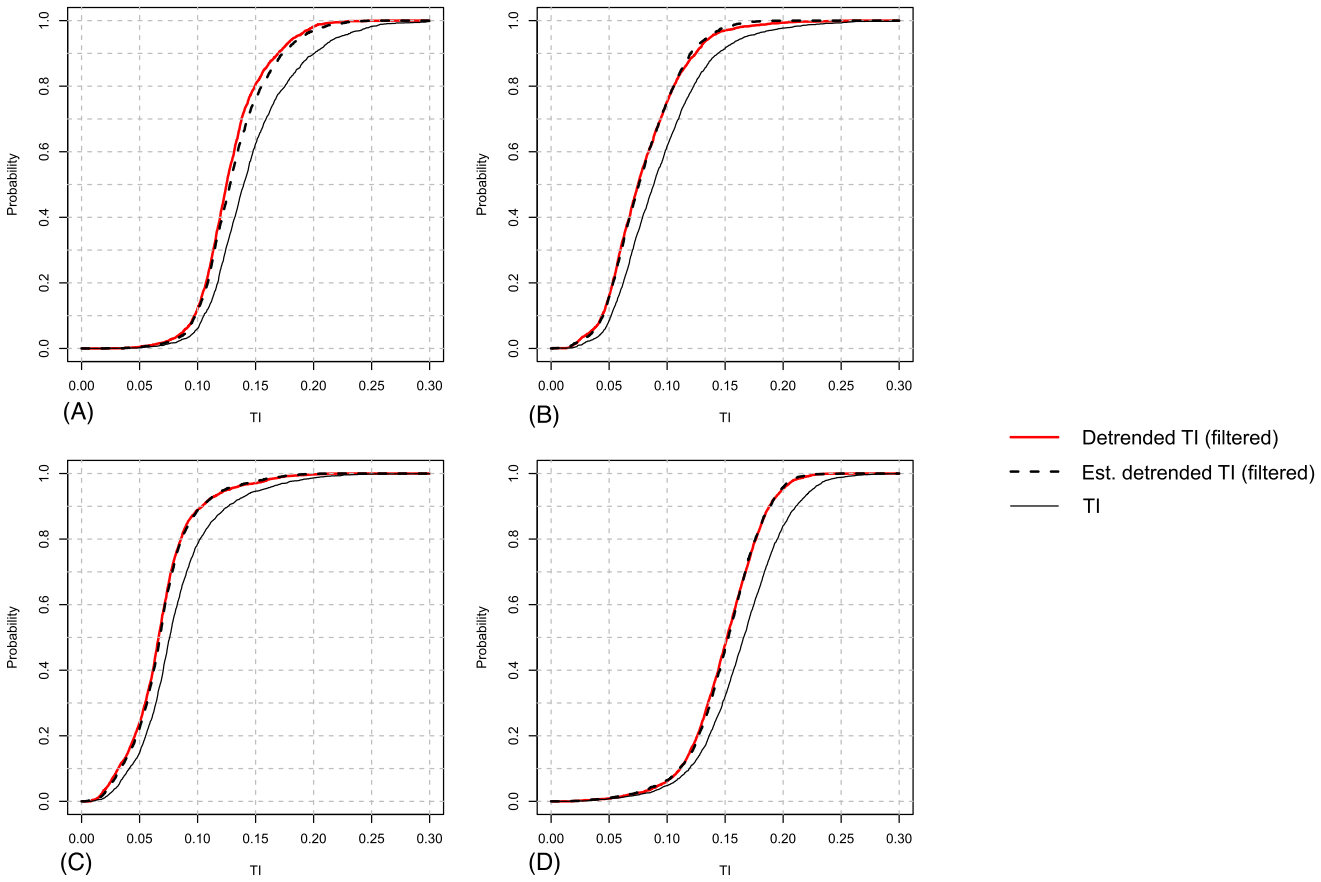

FIGURE 6 Cumulative distributions of TI in the filtered de-trending case. Estimated TIs are those resulting from the best performing, NN, model. Results from 4 sites are shown (A) Hanford (B) Sletringen (C) Orkney (D) Toboel

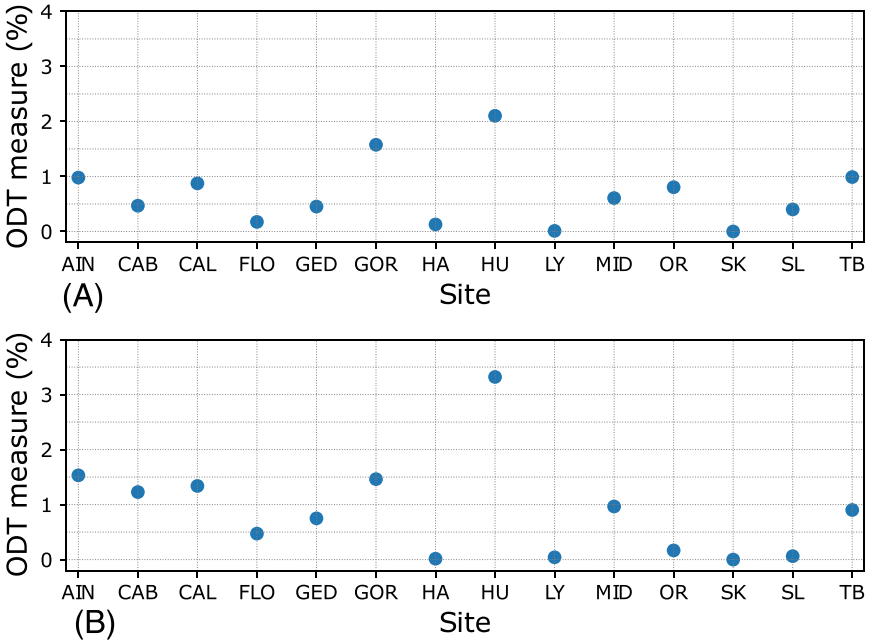

FIGURE 7 Over-de-trending measure values across test sites for (A) linear de-trending and (B) filtered de-trending

data. Potential benefits of a more complete treatment of windfield complexities when undertaking these analyses, including the application of robust filtering approaches, should therefore be explored. In addition, it seems sensible to expect turbine structural frequencies and dynamics to influence what constitutes optimal de-trending in any given case. The fatigue, hence frequency-domain, context of this problem implies filtering (of some kind) will likely remain the best approach to defining/identifying trends. The key question going forward is therefore likely to be what constitutes an optimal separation of turbulence and trend in terms of simulation based fatigue life prediction, and will optimal separation change as a function of turbine structural dynamics or site characteristics/conditions?

Two additional practicalities are worth considering with regards to the de-trending problem. The first is that a de-trended wind speed signal, while necessary for fatigue estimates based on stationary turbulence simulations, could lead to erroneous results if used as input to other analyses, such as extreme design load investigations.[10] The use, or not, of de-trended values is therefore context specific. Finally, while in the current work wind speed measurement frequencies were found to not be important as input variables for de-trended TI predictions, the study dataset was not large enough for this observation to be considered conclusive. Therefore, the possible role(s) of sampling frequency as it relates to wind measurements and accurate turbine fatigue assessment should remain a consideration in future work. In particular, while this work has been principally concerned with variance over-estimation, it should be appreciated that overly coarse sampling rates could introduce some amount of variance underestimation.

9 | CONCLUSIONS

This paper proposes a reformulation of moment only de-trending, from a trend identification problem, into a direct regression problem. In this context, available data quantities relate directly to the accuracy of resulting predictions, while the method itself is general, and so applicable for any method of de-trending. In the current work, both linear and filtered de-trending methods were considered. For each case, de-trended TI was modelled as a function of concurrent (plus five leads and lags) values of TI, mean wind speed and wind speed standard deviation, along with terrain type. A range of machine learning techniques were tested on data from across 14 sites in a 'leave-one-site-out' cross-validation procedure. Results demonstrated that machine learning based wind site moment de-trending can provide excellent approximations of the true de-trended turbulence distributions. Based on the presented results, recommended methods for de-trending in this context are Neural Networks, Random Forest and Boosted Trees. The Neural Network model resulted in the best performance overall, with improvements in RMSE over the currently recommended method of 13–45% in the case of linear de-trending. Finally, it was recommended that future work focus on identifying optimal methods and levels of de-trending for providing accurate remaining-life predictions, with answers potentially dependent on turbine structural dynamics and site conditions.

ACKNOWLEDGEMENTS

The authors would like to thank Julian Feuchtwang and Jim Brasseur for helpful discussions and suggestions during the course of this work. We would also like to thank the two anonymous reviewers for their valuable input.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1002/we.2720.

ORCID

Edward Hart https://orcid.org/0000-0002-2322-4520

REFERENCES

1. Topham E, McMillan D. Sustainable decommissioning of an offshore wind farm. *Renew Energy*. 2017;102:470-480.
2. Ziegler L, Gonzalez E, Rubert T, Smolka U, Melero J. Lifetime extension of onshore wind turbines: A review covering germany, spain, denmark, and the uk. *Renew Sustain Energy Rev*. 2018;82:1261-1271.
3. Hou P, Enevoldsen P, Hu W, Chen C, Chen Z. Offshore wind farm repowering optimization. *Appl Energy*. 2017;208:834-844.
4. Caughey SJ. Observed characteristics of the atmospheric boundary layer. *Atmospheric turbulence and air pollution modelling*. Dordrecht: Springer; 1984: 107-158.
5. Smedman A-S. Observations of a multi-level turbulence structure in a very stable atmospheric boundary layer. *Bound-Layer Meteorol*. 1988;44(3): 231-253.
6. Vickers D, Mahrt L. The cospectral gap and turbulent flux calculations. *J Atmos Ocean Technol*. 2003;20(5):660-672.
7. Basu S, Porté-Agel F, Foufoula-Georgiou E, Vinuesa J-F, Pahlow M. Revisiting the local scaling hypothesis in stably stratified atmospheric boundary-layer turbulence: an integration of field and laboratory measurements with large-eddy simulations. *Bound-Layer Meteorol*. 2006;119(3):473-500.
8. Larsen GC, Hansen KS. De-trending of wind speed variance based on first-order and second-order statistical moments only. *Wind Energ*. 2014;17: 1905-1924.

9. Hart E, Guy C, Tough F, Infield D. Wind site turbulence de-trending using statistical moments: evaluating existing methods and introducing a gaussian process regression approach. *Wind Energy*. 2021;24(9):1013–1030.
10. Hannesdóttir A, Kelly M, Dimitrov N. Extreme wind fluctuations: joint statistics, extreme turbulence, and impact on wind turbine loads. *Wind Energy Sci*. 2019;4(2):325-342.
11. IEC 61400-1:2019. Wind energy generation systems—part 1: design requirements. *Standard*, Geneva, Switzerland, International Electrotechnical Commission; 2019.
12. IEC 61400-13:2015. Wind turbines part 13: measurement of mechanical loads. *Standard*, Geneva, Switzerland, International Electrotechnical Commission; 2015.
13. Galinos C, Larsen TJ, Mirzaei M. Impact on wind turbine loads from different down regulation control strategies. *J Phys: Confer Ser*. 2018;1104: 012019.
14. Lio WH, Mirzaei M, Larsen GC. On wind turbine down-regulation control strategies and rotor speed set-point. *J Phys: Confer Ser*. 2018;1037:032040.
15. Hansen KS, Larsen GC. De-trending of turbulence measurements. In: European Wind Energy Conference & Exhibition. European Wind Energy Association (EWEA); 2007. Conference date: 07-05-2007 through 10-05-2007.
16. Damaschke M, Monnich K, Soker H. De-trending turbulence measurements identification of trends and their suppression. DEWI, http://www.windrose.gr/De-trending%20of%20TI%20(DEWI).pdf; 2007.
17. Robinson C, Schumacker RE. Interaction effects: centering, variance inflation factor, and interpretation issues. *Mult Linear Regression Viewpoints*. 2009; 35(1):6-11.
18. Afshartous D, Preston RA. Key results of interaction models with centering. *J Stat Edu*. 2011;19(3):1–24.
19. Lai J-P, Chang Y-M, Chen C-H, Pai P-F. A survey of machine learning models in renewable energy predictions. *Appl Sci*. 2020;10(17):5975.
20. Jørgensen KL, Shaker HR. Wind power forecasting using machine learning: State of the art, trends and challenges. In: IEEE 8th International Conference on Smart Energy Grid Engineering (SEGE); 2020:44-50.
21. Dimitrov N, Kelly MC, Vignaroli A, Berg J. From wind to loads: wind turbine site-specific load estimation with surrogate models trained on high-fidelity load databases. *Wind Energy Sci*. 2018;3(2):767-790. https://wes.copernicus.org/articles/3/767/2018/
22. James G, Witten D, Hastie T, Tibshirani R. *An Introduction to Statistical Learning: With Applications in R*. New York: Springer; 2013.
23. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning, Second Edition*. New York: Springer; 2009.

How to cite this article: Tough F, Hart E. Retrospective de-trending of wind site turbulence using machine learning. *Wind Energy*. 2022; 1-15. doi:10.1002/we.2720

APPENDIX A: QUANTIFYING TI REDUCTIONS—STATISTICAL ANALYSIS IN DETAIL

The average percentage *reduction* in TI was calculated across sites using high-frequency measurements, where

$$reduction = \frac{\overline{\mathrm{TI}} - \overline{\mathrm{TI}}^{\mathrm{dt}}}{\overline{\mathrm{TI}}} = 1 - \frac{\overline{\mathrm{TI}}^{\mathrm{dt}}}{\overline{\mathrm{TI}}}.$$

$\overline{\mathrm{TI}}$ represents average TI, and $\overline{\mathrm{TI}}^{\mathrm{dt}}$ represents average $\mathrm{TI}^{\mathrm{dt}}$. The distribution of *reduction* is bounded by 0 since reductions cannot be negative. Furthermore, the distribution of mean TI is positively skewed. Therefore, prior to estimation of the mean and standard deviation of the distribution, Box-Cox power transformations were used to transform the data to normality. The optimal Box-Cox power was $\lambda = 0.50$ for both linear and filtered de-trending. For linear de-trending, the mean reduction on the Box-Cox scale was $\mu_{\mathrm{reduc}^{(\lambda)}} = -1.46$ and standard deviation reduction $\sigma_{\mathrm{reduc}^{(\lambda)}} = 0.11$. For filtered de-trending, the mean reduction on the Box-Cox scale was equal to -1.31 with standard deviation equal to 0.11. On the Box-Cox scale, 95% prediction intervals are given by

$$\mu_{\mathrm{reduc}^{(\lambda)}} \pm t\sigma_{\mathrm{reduc}^{(\lambda)}}\sqrt{1+\frac{1}{n}},$$

where $t = 2.16$. Back transformation to the original scale then results in the intervals presented in Section 3. These intervals incorporate uncertainty of the mean via standard error of the mean and uncertainty in standard deviation using a t distribution. The Shapiro-Wilk test was used to assess approximate normality on the Box-Cox scale. As the datasets span a range of regions, seasons and measurement frequencies, these 95% intervals are generalised. Too few observations were available to regress the parameters as functions of covariates; however, it is expected that distributions could be broken down in relation to input variables given a larger study database.