RESEARCH ARTICLE

Wind site turbulence de-trending using statistical moments: Evaluating existing methods and introducing a Gaussian process regression approach

Edward Hart[1] 🟢 | Callum Guy[2] | Fraser Tough[3] | David Infield[1]

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

[2]Institute for Energy Systems, School of Engineering, University of Edinburgh, Edinburgh, UK

[3]Renewable Energy Systems Ltd., Glasgow, UK

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

Funding information
EPSRC, Grant/Award Numbers: EP/R513349/1, EP/L016680/1

Abstract

This paper considers the problem of retrospectively de-trending wind site data when only statistical moments, in the form of 10-min means and standard deviations in wind speed, are available. Low-frequency trends present in wind speed data are known to bias fatigue damage estimates, and, hence, removal of their influence is important for accurate fatigue life estimation. When raw data is available, this procedure is straightforward; however, for many sites, significant quantities of data are available, which contain only statistical moments. Additional value is therefore unlocked if de-trending can also take place in this context. Existing methods, Models 1 and 2, are introduced, and their performance and viability appraised, respectively. A Gaussian process (GP) regression implementation is also developed, which seeks to incorporate characteristics of real trends extracted from raw data into the fitting procedure via an appropriately chosen lengthscale hyperparameter. Results indicate that Model 2, the recommended method in previous work, suffers from fundamental issues, with the implication that it should no longer be used. Model 1 and GP results are shown to be very similar at the turbulence distribution level. This finding is interpreted as a validation of Model 1 and an indication that it may already be performing as well as can be hoped for, given the information available in the current problem formulation. Theoretical overheads associated with GPs, in addition to the performance similarities mentioned above, lead to Model 1 being recommended as the best approach to moment-based turbulence de-trending at this time.

KEYWORDS
de-trending, Gaussian process, site conditions, turbulence, wind energy

1 | INTRODUCTION

Statistical assessment of site wind resource has important implications for wind farm performance aside from just expected energy yield. In terms of wind turbine structural fatigue, site turbulence intensity (TI) distributions play a key role in the prediction of fatigue life consumption. This in

---

turn has implications for design and lifetime-extension decisions for wind farms and, hence, is an economically important area for wind farm owners and operators.[1-3]

When wind speed statistics are calculated for a wind farm site, typically as 10-min means and standard deviations (which in turn are used to calculate TI values), an implicit assumption of stationarity is being applied. However, wind speed variations occur across a wide range of frequencies, and so over a given capture window, the standard deviations calculated contain contributions from components slowly varying within the chosen 10-min window. This is known to bias results and can lead to significant errors in fatigue-life estimation.[1] Therefore, for fatigue considerations, one should ideally separate wind speed variance into 'macro-scale' and 'micro-scale' contributions, with macro-scale contributions treated as a *trend* over each capture window. Explicitly, the technical definition of *trend* adopted in the current work is *the summation of all Fourier modes present in a wind speed time series with period longer than the capture window*. This definition coincides with that of Larsen and Hansen.[1] For the case where raw data are available at a site, there are various methods which exist with which different levels of de-trending can be obtained, the most common being a straight line fit to the data. However, at many sites, years worth of resource or operational data are available for which only 10-min first- and second-order moments (i.e., means and standard deviations) are available. For such data, these available de-trending approaches cannot be undertaken. Therefore, techniques which can effectively de-trend using only 10-min moments, and so retrospectively remove the influence of low frequency contributions to turbulence, have the potential to significantly improve the accuracy of fatigue life estimates and contribute to life-extension efforts. This moment-based de-trending problem forms the focus of this current study, with the seminal work of Larsen and Hansen[1] taken as a starting point. Their two approaches both facilitate the fitting of splines to moment data in order to recreate possible trend contributions. Given that spline orders effectively codify a proportion of smoothing when fitting, it seems pertinent to ask what the correct amount of smoothing might be in this context. This question is considered through the development of a Gaussian process (GP) machine learning approach to moment-based de-trending. A lengthscale parameter within the GP's covariance structure defines the quantity of smoothing during regression. An appropriate lengthscale is sought using available data, allowing comparisons to be made between existing methods and one seeking to apply a level of smoothing which reflects trend component characteristics in real wind site data.

Section 2 describes relevant background material including notation, problem definition and performance requirements. Section 3 then introduces existing approaches to moment-based de-trending and considers the performance and viability of these methods. In Section 4, a GP regression approach to the problem is developed, with results and comparisons given in Section 5. Conclusions and recommendations then close the paper.

2 | BACKGROUND

In this section, the problem and notation will be presented in detail, along with relevant background material and a discussion of performance requirements and performance measures.

2.1 | Notation and problem definition

Given the varied notations used across the sources and fields from which this study draws, the one outlined below has been chosen to enhance clarity and avoid confusion with the later defined GP notation. Letting τ denote the time between subsequent measurements, and assuming $T/\tau = N \in \mathbb{N}$, let

$$i\underline{T} = \{iT + n\tau\}_{n=0}^{N}, \tag{1}$$

for a given integer $i \geq 0$. These sets, $i\underline{T}$, therefore contain the time steps of the ith bin. Wind speeds at each time step, v_t for a given $t \in i\underline{T}$, are assumed to have been measured; throughout the paper, these underlying measurements will be referred to as *raw data*. From this, the first two statistical moments over each bin are calculated, these being the mean,[*]

$$M_{i\underline{T}} = \frac{1}{N} \sum_{t' \in i\underline{T}} v_{t'}, \tag{2}$$

and variance,

$$V_{i\underline{T}} = \frac{1}{N} \sum_{t' \in i\underline{T}} \left(v_{t'} - M_{i\underline{T}} \right)^2, \tag{3}$$

---

[*]Similar to the use of dummy variables within integrals, summation variables here are denoted as either t' or t^*. t itself will be used to denote specific, individual time steps.

referred to collectively as *moment data* throughout the paper and from which TI is calculated as

$$TI_{i\underline{T}} = \frac{\sqrt{V_{i\underline{T}}}}{M_{i\underline{T}}}.$$ (4)

It is assumed that wind speed measurements are composed of a low-frequency trend, g_t, and higher frequency stochastics, s_t, such that

$$v_t = g_t + s_t,$$ (5)

with g_t consisting of all Fourier modes of period greater than T as per the trend definition given in Section 1. Assuming independence of the two components, their variances decompose additively,

$$V_{i\underline{T}} = V_{i\underline{T}}^g + V_{i\underline{T}}^s,$$ (6)

where

$$M_{i\underline{T}}^g = \frac{1}{N} \sum_{t' \in i\underline{T}} g_{t'},$$ (7)

$$V_{i\underline{T}}^g = \frac{1}{N} \sum_{t' \in i\underline{T}} \left(g_{t'} - M_{i\underline{T}}^g \right)^2,$$ (8)

and similarly for s. In the context of TI de-trending, the current problem becomes as follows:

From measurements of $M_{i\underline{T}}$ and $V_{i\underline{T}}$, is it possible to estimate the stochastic variance component, $V_{i\underline{T}}^s$, such that the standard deviation,

$$\sqrt{V_{i\underline{T}}^s},$$

can be used to form de-trended values of TI which are unaffected by low-frequency trend contributions?

The methods considered in this work approach this by seeking to identify the low-frequency trend, g_t, via curve fitting of one kind or another. From there, $V_{i\underline{T}}^g$ can be computed directly, and it follows that the sought variance is obtained from measured $V_{i\underline{T}}$ as

$$V_{i\underline{T}}^s = V_{i\underline{T}} - V_{i\underline{T}}^g.$$ (9)

2.2 | Linear and nonlinear de-trending

It is important to note the formal definition of a trend as presented in Section 1. From this, it follows that if one has access to the non-moment data, i.e., the underlying measurements themselves, then trend identification involves a simple application of low-pass filtering with a cutoff frequency corresponding to the width of capture bin (1/600 Hz for 10-min bins). As such, the trend identification problem when underlying data are available is considered well understood, and such filter-identified trends are used in the current work as a baseline for performance comparisons on moment data. A third-order Butterworth low-pass filter was implemented for this purpose; code for such filters is standard and readily available.[4] An example filtered trend is shown against underlying data over six 10-min bins in Figure 1. Subtracting the identified trend from underlying data allows the de-trended TI distribution to be calculated. Within the wind industry, when underlying data are available, it is still very common to perform straight line fits within each window[5] (rather than following a filtering approach) and to treat this line as the trend. The existing methods for moment de-trending are therefore geared towards subtracting the linear trend portion of variance values. The differences from following a linear or filtering (i.e., nonlinear) approach to de-trending when underlying data is available are clear in Figure 1. In general, this same pattern holds, and, overall, linear de-trending gives a conservative reduction in TI compared with the filtering approach—indicating there is an opportunity in general to improve over linear fits for de-trending purposes. However, such considerations do not form the main focus of this paper and so are not considered further. With respect to the moment de-trending problem, the relevance of considering both linear- and filter-based de-trending is twofold: (1) to provide an overall baseline for the most variance one can hope to remove from a given TI distribution. (2) Whereas existing methods are geared towards removing the linear trend portion, they are not limited to this; furthermore, the GP method only considers nonlinear fits. It is therefore concluded that a principled comparison of such techniques requires for performance to be measured against both the linear- and filter-based approaches to de-trending wind field data.
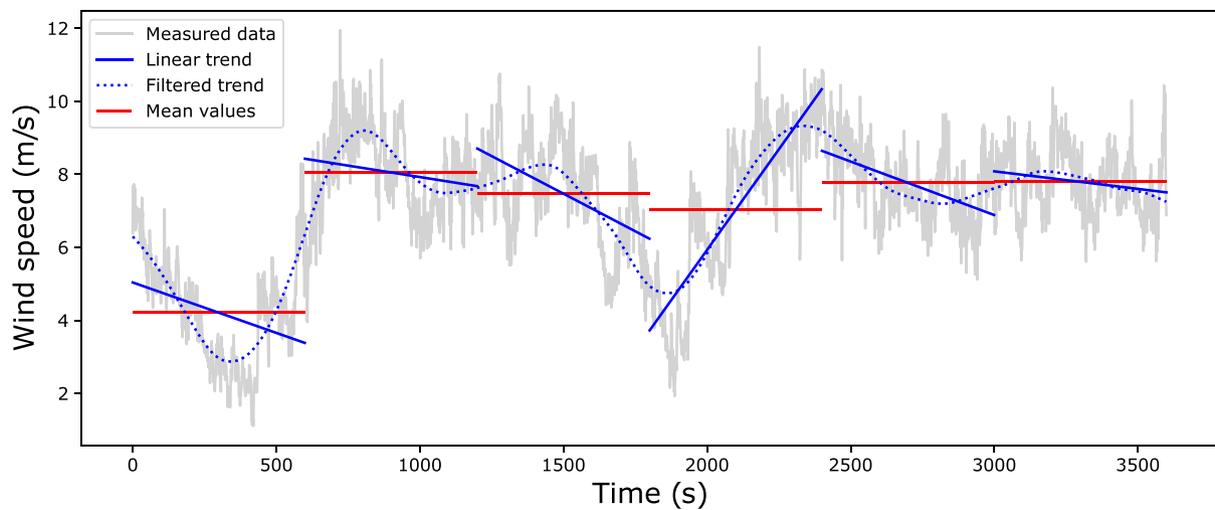
FIGURE 1 Raw wind speed data over a 1-h duration, along with mean values and linear trend fits across each 10-min (600 s) bin. The low-frequency trend obtained using low-pass filtering with a cutoff frequency of 1/600 Hz is also shown [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 1 Wind site datasets—Summary information

Site	Terrain	Measurement height (m)	Usable data (hours)
Tobøl, DK	Pastoral	64	607
Oak Creek, USA	Complex	79	525
Hanford, USA	Complex	40	276
Skipheia, NO	Coastal	45	2,587
Sletringen, NO	Coastal	45	3,691
Orkney, UK	Coastal	64	6,076
Horns Rev, DK	Offshore	50	525
Middelgrunden, DK	Offshore	45	1,513

2.3 | Wind site datasets

Summary information for wind site data used in this work is presented in Table 1.[†] Note that the *Usable Data* value given in Table 1 for each site is that available after the removal of any 10-min bins in which over 20% of datapoints are missing. This criterion was applied to avoid instances of significant missing data within individual bins artificially influencing results. Higher-frequency measurements were downsampled to 1 Hz to maintain reasonable computational times and memory efficiency for this comparative study.

2.4 | Performance measures

For any method, it is important to consider the context in which it is to be applied when determining how to assess performance. As outlined in Section 1 and discussed in detail in Larsen and Hansen,[1] the current problem of turbulence de-trending has the potential to influence both design and life-extension decisions for entire wind farms, with both areas having significant cost impacts to wind projects. In this context, it should be clear that conservatism must be favoured above over-optimism, with over de-trending (ODT; and so an under-estimation of site turbulence levels) having potentially catastrophic implications for project economics. When assessing performance in this case, the best available method will therefore be that which provides the highest level of de-trending while avoiding any significant quantity of ODT (when comparing the distribution obtained from moment data to baseline distributions generated from de-trending on underlying raw data). Explicitly, since a distribution shifted too far left must be regarded differently from the same distribution shifted an identical distance right, best fits between computed and baseline distributions must in this case be considered in an asymmetrical manner.

[†]This data was obtained from https://www.winddata.com/

These aspects of performance are captured using the two performance metrics illustrated in Figure 2. The first is an ODT measure between computed (co) and baseline (bl) distributions,

$$E_{\mathrm{ODT}}(p_{\mathrm{co}}, p_{\mathrm{bl}}) = \int_{-\infty}^{x_{\mathrm{Mo}}} \max(p_{\mathrm{co}}(x) - p_{\mathrm{bl}}(x), 0) dx, \tag{10}$$

where x_{Mo} is the value of x which corresponds to the *mode* (i.e., maximum) of the baseline distribution, p_{bl}. E_{ODT} expresses the probability mass in the computed distribution which lies to the left of the baseline, thus capturing the total quantity of over de-trending present. The second is the Wasserstein or Earth-Movers Distance (EMD),[6,7] which captures the probabilistic *work done* in transforming the original (or) distribution into the de-trended (dt) distribution.

$$E_{\mathrm{EMD}}(p_{\mathrm{dt}}, p_{\mathrm{or}}) = \min_{\pi \in \mathcal{P}(p_{\mathrm{dt}}, p_{\mathrm{or}})} \int_{\mathbb{R} \times \mathbb{R}} |x - x'| d\pi, \tag{11}$$

for $(x, x') \sim \pi$ and where $\mathcal{P}(p_{\mathrm{dt}}, p_{\mathrm{or}})$ is the set of all joint probability distributions on $\mathbb{R} \times \mathbb{R}$ whose marginal distributions have densities p_{dt} and p_{or} in the first and second dimensions, respectively. E_{EMD} is therefore a measure which describes the overall magnitude of de-trending taking place. With respect to Figure 2B, the depicted infinitesimal area (Δweight) corresponds to $d\pi$ and the distance (d) to $|x - x'|$, with their product (ΔWork) contributing to the overall integral of work done. Note that, throughout the paper, *computed* distributions are those generated by applying a given moment de-trending method, *baseline* distributions are those obtained via linear or filtered de-trending on raw data and *original* refers to unaltered TI distributions on raw data.

Together, the ODT and EMD measures allow for performance assessment of the type required in the current context, with successful de-trending indicated by an E_{ODT} of acceptably low level, alongside a high value of E_{EMD}. Importantly, any method resulting in E_{ODT} values above what is deemed a reasonable level must necessarily be considered unsound, no matter how large its corresponding E_{EMD} or how close this EMD value is to that of the baseline distribution (de-trended using raw data) against which comparisons are being made. This latter consideration is where the required asymmetry, regarding over and under de-trending, manifests.

3 | LARSEN AND HANSEN MOMENT DE-TRENDING MODELS

The moment de-trending methods developed in Larsen and Hansen[1] will now be presented. These methods consist of two spline models, referred to as Model 1 (M1) and Model 2 (M2), which look to fit fourth- and third-order polynomials, respectively, within each 10-min bin subject to a range of continuity conditions. M1 fits to mean values alone, whereas M2 attempts to leverage variance information as well within the fitting procedure. In each case, the linear trend contribution to be subtracted in the ith bin takes the form
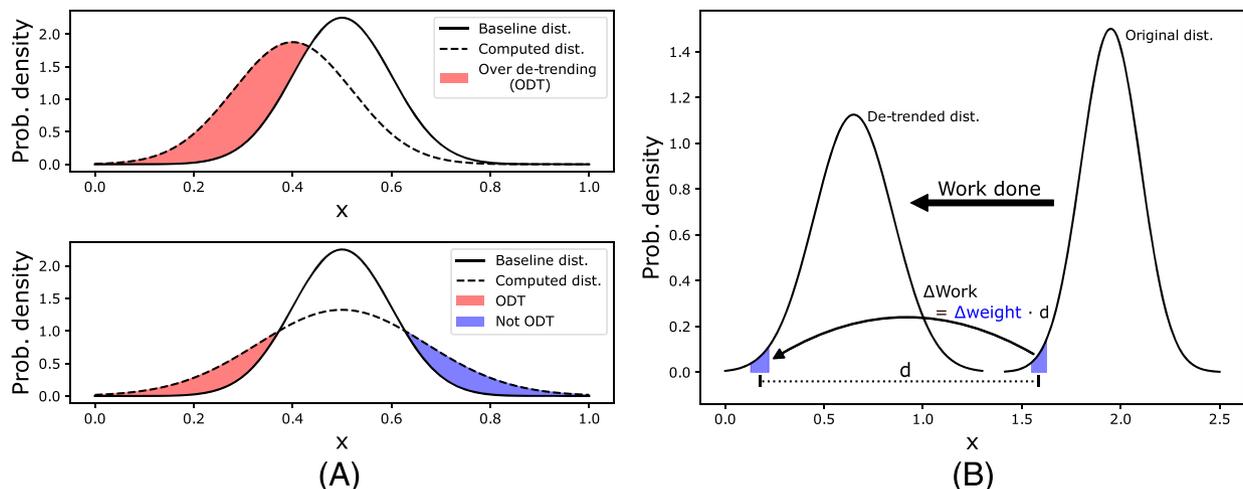
FIGURE 2 Graphical representations of performance measures used to capture (A) over-detrending and (B) probabilistic work done (Earth-Movers distance) [Colour figure can be viewed at wileyonlinelibrary.com]

$$\frac{T^2}{12}h_i^2, \tag{12}$$

where h_i (different for each of M1 and M2) is the gradient of the straight line fit to the resulting polynomial in the ith bin—a linear function in fitted polynomial coefficients. M2 makes direct use of h_i terms when solving, and so is inherently tied to the linear de-trending case. M1 does not, and so one has the choice as to whether to remove just the linear portion of trend, or the full variance of fitted polynomials.

Each model can theoretically be fitted simultaneously across an arbitrary number of bins, although it should be noted that M2 makes the explicit assumption that the component of TI unrelated to the trend remains constant across the set of bins fitted to. In the original work,[1] the existence of a spectral energy gap,[8] across time periods ranging from 10 min to over an hour, is invoked to select three bins (30 mins) as the number to perform fitting on simultaneously. For M2, where a common TI value is solved for and explicitly applied across all bins being fitted to, this logic certainly holds. Indeed, testing on larger numbers of bins revealed a degradation in performance likely due to this consideration. For M1 and the later developed GP approach however, such common parameters are not part of the fitting process, and hence, it was deemed possible to increase the number of bins fitted up to six (1 h). This means that fitting is still performed across timescales associated with the spectral gap, but in this case, more information is made available to the fitting procedure at one time. In practice, the application of M1 to three or six bins was found to result in negligible differences. Results throughout the paper correspond to M1 (and GP) on six bins and M2 on three bins. The following sections introduce M1 and M2 in detail, assuming the specified number of bins is fitted to in each case.

3.1 | Model 1

In M1, the trend component within each bin takes the form of a fourth-order polynomial,

$$p_i(t) = a_i t^4 + b_i t^3 + c_i t^2 + d_i t + e_i, \text{ for } t \in i\underline{T} \quad \text{and} \quad i = 0, \ldots, 5, \tag{13}$$

subject to the following conditions:

$$
\begin{aligned}
\frac{1}{T}\int_{t^* \in i\underline{T}} p_i(t^*)\,dt^* &= M_{i\underline{T}} \text{ for } i = 0, \ldots, 5 &&(\textit{function means must equal window means}), \\
p_i(iT) - p_{i-1}(iT) &= 0 \text{ for } i = 1, \ldots, 5 &&(\textit{continuity at window intersections}), \\
p_i'(iT) - p_{i-1}'(iT) &= 0 \text{ for } i = 1, \ldots, 5 &&(\textit{first derivative continuity}), \\
p_i''(iT) - p_{i-1}''(iT) &= 0 \text{ for } i = 1, \ldots, 5 &&(\textit{second derivative continuity}), \\
p_i'''(iT) - p_{i-1}'''(iT) &= 0 \text{ for } i = 1, \ldots, 5 &&(\textit{third derivative continuity}), \\
p_0''(0) &= 0 &&(\textit{boundary condition for system closure}), \\
p_0'''(0) &= 0 &&\text{''}, \\
p_5''(6T) &= 0 &&\text{''}, \\
p_5'''(6T) &= 0 &&\text{''}.
\end{aligned}
\tag{14}
$$

Each of these can be resolved into a linear expression in the polynomial coefficients for each value of i. For example, the first conditions become[‡]

$$\left(\frac{1}{5T}\left[t^5\right]_{iT}^{(i+1)T}\right)a_i + \left(\frac{1}{4T}\left[t^4\right]_{iT}^{(i+1)T}\right)b_i + \left(\frac{1}{3T}\left[t^3\right]_{iT}^{(i+1)T}\right)c_i + \left(\frac{1}{2T}\left[t^2\right]_{iT}^{(i+1)T}\right)d_i + e_i = M_{i\underline{T}}, \tag{15}$$

for each i, and similarly, the fifth (third derivative) conditions become

$$(24iT)a_i + 6b_i - (24iT)a_{i-1} - 6b_{i-1} = 0, \tag{16}$$

for each i. Taken together, this constitutes a linear system of 30 equations and 30 unknowns—the polynomial coefficients across all bins being fitted to. Hence, the system can be expressed in the form

$$A\mathbf{c} = \mathbf{b}, \tag{17}$$

[‡]Note $[y(t)]_{t_1}^{t_2} = y(t_2) - y(t_1)$.

with A populated by the coefficient multipliers in the system, $\mathbf{c}$ the polynomial coefficients for all values of i and $\mathbf{b}$ the system's right-hand side values (bin means and zeros). The coefficients which solve this system are therefore

$$\mathbf{c} = A^{-1}\mathbf{b}. \tag{18}$$

3.2 | Model 2

In M2, the trend component within each bin takes the form of a third-order polynomial,

$$p_i(t) = a_i t^3 + b_i t^2 + c_i t + d_i, \text{ for } t \in i\underline{T} \text{ and } i = 0, 1, 2, \tag{19}$$

subject to the following conditions:

$$
\begin{aligned}
\frac{1}{T} \int_{t^* \in i\underline{T}} p_i(t^*) dt^* \quad &= M_{i\underline{T}} \text{ for } i = 0, 1, 2 \quad && \text{(function means must equal window means)}, \\
p_i(iT) - p_{i-1}(iT) \quad &= 0 \text{ for } i = 1, 2 \quad && \text{(continuity at window intersections)}, \\
p_i'(iT) - p_{i-1}'(iT) \quad &= 0 \text{ for } i = 1, 2 \quad && \text{(first derivative continuity)}, \\
p_i''(iT) - p_{i-1}''(iT) \quad &= 0 \text{ for } i = 1, 2 \quad && \text{(second derivative continuity)}, \\
V_{i\underline{T}} = \Gamma(M_{i\underline{T}}) + \frac{T^2}{12} h_i^2 \text{ for } i = 0, 1, 2 \quad && \quad && \text{(variance components sum to measured variance)}, \\
p_0''(0) \quad &= 0 \quad && \text{(boundary condition)}, \\
p_2''(3T) \quad &= 0 \quad && \text{"}.
\end{aligned}
\tag{20}
$$

Since the fifth condition contains an implicit (and nonlinear) term, iteration is required in order for solutions to be found. Based on considerations of offshore and onshore conditions, the function Γ takes on a different form in each case. For *onshore conditions*, where surface roughness is fixed,

$$\Gamma(M_{i\underline{T}}) = C^2 M_{i\underline{T}}^2, \tag{21}$$

with[§]

$$C = \frac{0.99}{3} \sum_{i=0}^{2} C_{-1}(a_i, b_i, c_i, d_i) + \frac{0.01}{3} \sum_{i=0}^{2} C_{-2}(a_i, b_i, c_i, d_i), \tag{22}$$

where subscripts -1 and -2 denote use of the most recent and second most recent estimates of the polynomial coefficients, respectively, across the three bins from previous iterations and for $* = -1$ or -2,

$$C_*(a_i, b_i, c_i, d_i) = \frac{1}{M_{i\underline{T}}} \sqrt{V_{i\underline{T}} - \frac{T^2}{12} h_i^2(a_i, b_i, c_i, d_i)}. \tag{23}$$

These values can be seen to be estimated from the fifth condition in Equation 20.

For *offshore conditions*, where surface roughness is related to wind speed,

$$\Gamma(M_{i\underline{T}}) = (2.39)^2 \left(\left[r + \sqrt{q^3 + r^2} \right]^{1/3} + \left[r - \sqrt{q^3 + r^2} \right]^{1/3} + \frac{4}{3} \right)^2, \tag{24}$$

with

$$q = \ln\left(\left(\frac{k_c}{gz} \right)^{-1/3} \right) - \frac{7}{9}, \tag{25}$$

[§]This next expression averages the most recent C values across the windows while also performing a relaxation of the transition from previously estimated values.

$$r = \ln\left(\left(\frac{k_c}{gz}\right)^{2/3}\right) + \frac{\kappa M_{iT}}{2} + \frac{10}{27}. \tag{26}$$

g is acceleration due to gravity, κ is the von Karman constant, z is the measuring altitude and k_c is the Charnock constant. This last parameter must be estimated, in a similar way to C in the onshore case, by again using previously estimated coefficients and relaxation within an iterative scheme.

For both the onshore and offshore cases, given initial (or the most recent) estimated coefficient values, the implicit terms in the equation set can be estimated, resulting again in a linear system of the form

$$A\mathbf{c} = \mathbf{b}. \tag{27}$$

The number of coefficients versus equations this time results in an overdetermined system and so a 'best-fit' approximate solution for the current iterate is found using the Moore-Penrose pseudo inverse, A^+, giving

$$\mathbf{c} = A^+\mathbf{b}. \tag{28}$$

From here, the estimates of implicit terms can be updated, and the system again solved as above to determine an updated set of coefficient solutions. This process is iterated in the hopes of converging to an overall solution. The recommended convergence criteria[1] considers the magnitude of coefficient and parameter value changes between iterations and stops the process once this chosen metric falls below a critical value.

3.3 | Evaluating performance of the Larsen and Hansen models

M1 and M2 were used to de-trend wind site data, with results then compared to those obtained via raw data linear- and filter-based de-trending approaches.

3.3.1 | Model 1 performance

M1 was found to behave very much as described in Larsen and Hansen.[1] Figure 3 shows example results for Hanford and Sletringen, alongside the original TI distributions and those resulting from the two types of raw data de-trending. Results for M1 from applying the linear de-trending formula or removal of the full fitted trend variance show negligible differences in resulting distributions; Figure 3 shows linear de-trending formulation results. Note that in this implementation, and those of the other methods in the current work, any de-trending which results in a negative
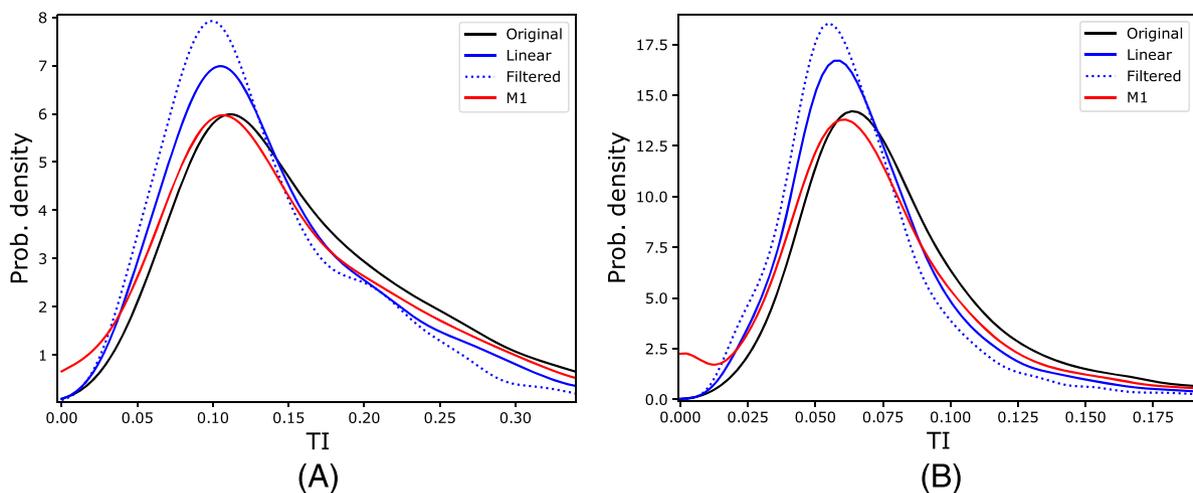
FIGURE 3 Moment de-trending using Model 1 for (A) Hanford and (B) Sletringen datasets. For comparison, the original distributions and those obtained from linear- and filter-based de-trending using raw data are also shown [Colour figure can be viewed at wileyonlinelibrary.com]

residual variance has its de-trended value instead set equal to zero (since variance cannot be negative). In such cases, the method has indicated it believes all measured variance is due to the trend component. Negative values occur rarely and generally over portions of wind speed data in which signal variance is exceedingly low, with negative values essentially indicating over-fitting (and hence over de-trending) on the part of the method, although over-fitting can also occur without resulting in negative values. Increases in probability density at zero TI, seen in M1 and GP results, are mostly driven by this handling of negative values since all such cases become concentrated at the origin.

M1 can be seen to manage some amount of de-trending while avoiding any significant tendency to over de-trend (the visible areas where this occurs accounting for no more than 3% of the overall distribution). In Section 5, where comprehensive results are presented for both M1 and the GP method together, it will be shown that these observations hold in general.

3.3.2 | The viability of M2

In Larsen and Hansen,[1] the terms within the vector $\mathbf{b}$ of Equation 28 which implicitly rely on the coefficient values being solved for are shown to be Γ_i and $\text{sign}(h_i)$ for $i = 0, 1, 2$. For the sake of clarity, the following analysis proceeds considering the onshore equations, although it can be seen that the offshore case will follow along identical lines. Hence, in this setting, it is C, constant across the three bins, and $\text{sign}(h_i)$ for $i = 0, 1, 2$ (collectively denoted by $\text{sign}(\mathbf{h})$) which together form the implicit terms in $\mathbf{b}$. At this stage, it is important to note the complexity of the problem at hand. With no exact solution available (due to having an over-determined system) and the presence of one continuous variable (C) and a set of eight possible sign combinations ($\text{sign}(h_i) = \pm 1$ for $i = 0, 1, 2$), finding a best-fit solution corresponds to determining the global minimum of the following error,

$$||A\mathbf{c} - \mathbf{b}(\mathbf{c})||_2, \tag{29}$$

as a function of C and $\text{sign}(\mathbf{h})$, since, $\mathbf{c} = A^+ \mathbf{b}(C, \text{sign}(\mathbf{h}))$. Note the deliberate difference in the independent variables attributed to $\mathbf{b}$ in this latter expression compared to Equation 29. This is due to the fact that once a possible vector of coefficients has been solved for using given values of C and $\text{sign}(\mathbf{h})$, the correct $\mathbf{b}$ vector against which to test error is necessarily the one formed from the newly generated coefficients.

In the original paper, the presented iterative method is used as a route to try and find a best-fit solution, with each subsequent estimate of polynomial coefficients used to determine the C and $\text{sign}(\mathbf{h})$ values of the next iterate, and with convergence assumed to indicate success in locating such a solution. However, viewing the problem as described above, it should be clear there is significant complexity in this cost function, with a continuous range of possible C values coupled to eight possible $\text{sign}(\mathbf{h})$ combinations. It therefore seems prudent to ask whether local, as well as a global, minima might be present and whether the iterative scheme is adversely effected by this. This question was investigated by implementing M2 in two different ways. Firstly, an iterative scheme identical to that presented in Larsen and Hansen[1] was used to de-trend site data. Secondly, for these same sites, a more computationally expensive *global optimisation* was performed whereby coefficients were solved for across 100 values of C between 0 and the maximum TI (this being what C represents) over the three bins considered in each window, and with each of the eight possible combinations of $\text{sign}(\mathbf{h})$ values. The error was calculated for each resulting vector of coefficients, as per Equation 29, with the coefficient vector corresponding to minimum overall error taken to be the global best-fit solution. Example TI distributions resulting from the *iterative-* and *optimisation*-based approaches to M2 moment de-trending are shown in Figure 4, along with the original distributions and those obtained from linear- and filter-based de-trending using raw data for comparison. It was found that the iterative approach behaved as described in the original paper, with the resulting de-trended distributions generally providing a reasonable fit to distributions obtained from linear de-trending using raw data. The optimisation approach on the other hand leads to very different results, with distributions clearly over de-trended when compared to both linear- and filter-based de-trending using raw data. The characteristics of the cost function for which these methods are looking to identify a minimum was therefore considered in more detail in order to understand this discrepancy. Figure 5 shows this cost function for two example windows across the explored range in C, and for each possible combination of $\text{sign}(\mathbf{h})$ values, the final value of C reached via iteration is also indicated. Note that in the optimisation implementation, only physically meaningful solutions are considered as viable,[¶] this being why some lines end prematurely in the figure. These two examples are typical of what is seen throughout the application of these methods to wind site data. It can be seen that these are indeed highly complex cost functions, with large numbers of local minima. Furthermore, the iterative method has in both cases converged to a local rather than global minimum, with the convergence also being within a sub-optimal $\text{sign}(\mathbf{h})$ combination in sub-figure (b). In the vast majority of observed cases, the value of C corresponding to the global minimum is found to lie to the left of that reached via the iterative implementation. Considering condition 5 of Equation 20, it follows that the value of C being smaller necessarily implies a larger variance contribution attributed to the trend, and so a greater magnitude of de-trending when this contribution is removed. This aligns with the TI distribution results, with significantly more de-trending taking place under the optimisation approach. Figure 6 shows the ODT and EMD results corresponding to the distributions in Figure 4 which further corroborate the observations made thus far. Whereas the over-detrending as

---

[¶]Some solutions can result in a trend contribution which contains more variance than the original measurement, which is clearly not physically possible.
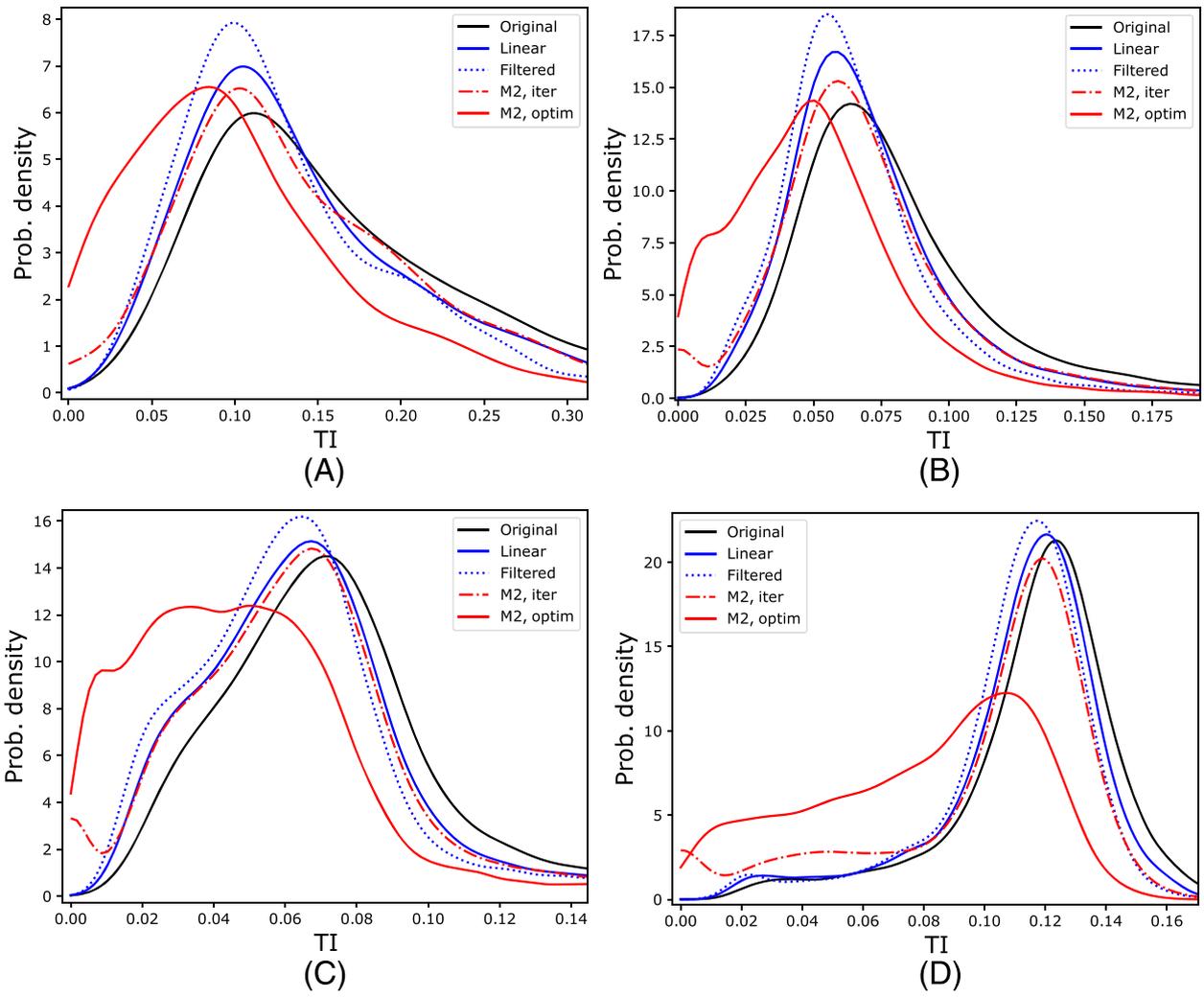
FIGURE 4 Moment de-trending using iterative- and optimisation-based approaches to Model 2 from Larsen and Hansen[1] for (A) Hanford , (B) Sletringen, (C) Orkney, and (D) Toboel datasets. For comparison, the original distributions and those obtained from linear- and filter-based de-trending using raw data are also shown [Colour figure can be viewed at wileyonlinelibrary.com]
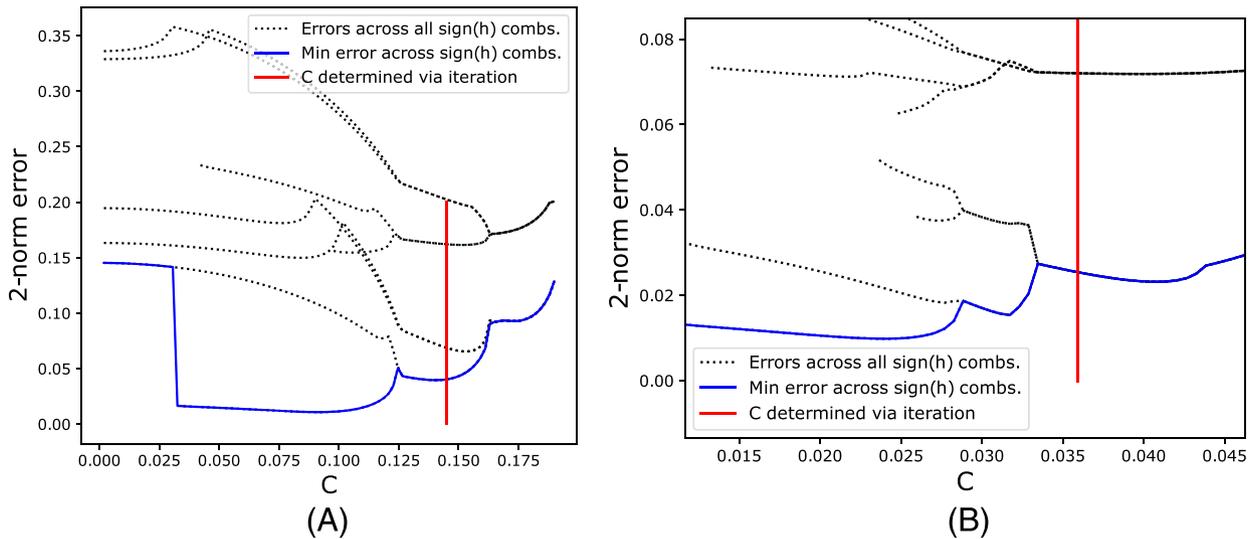
FIGURE 5 Cost function examples across the explored ranges in C, and for each possible combination of sign($\mathbf{h}$) values, the final value of C reached via iteration is also indicated [Colour figure can be viewed at wileyonlinelibrary.com]
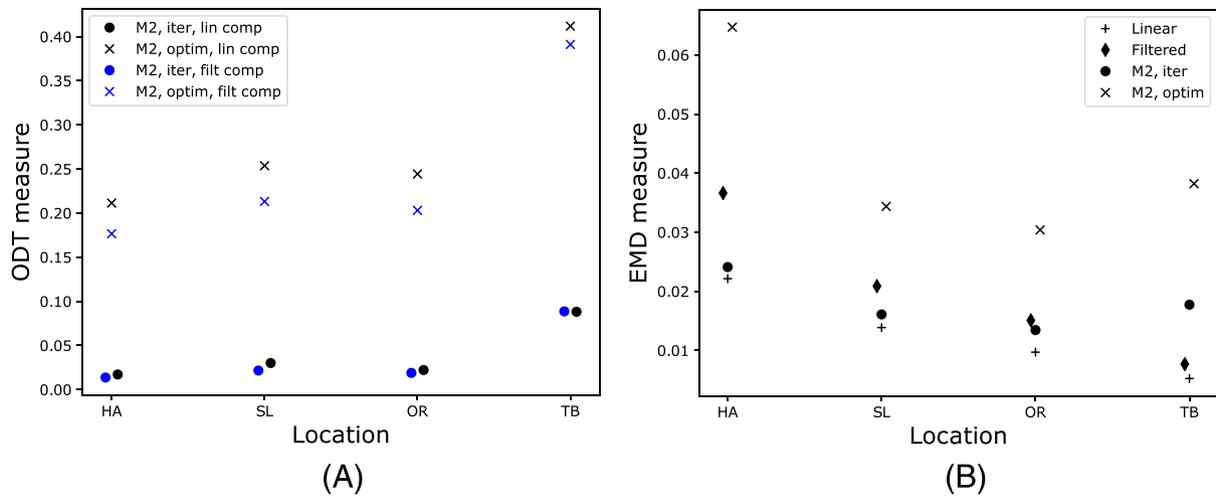
FIGURE 6 Performance results in the form of over de-trending (ODT) and earth-movers distance (EMD) measures for Model 2 in subplots (A) and (B), respectively. Results are shown for both iterative- and optimisation-based implementations [Colour figure can be viewed at wileyonlinelibrary.com]

measured for iterative M2 remains close to zero, reaching around 10% for only a single site, the true best-fit solutions (obtained via optimisation) instead have ODT values ranging from 10% to 40%. With respect to EMD values, the true best-fit solutions to the M2 equation system obtained from optimisation can be seen to result in a much greater amount of work done than for all other methods in general. Taken together, while bearing in mind the performance requirements discussed in Section 2.4, the above analyses and presented results necessarily lead to the following two conclusions:

1. The iterative approach cannot reliably locate the best fit solution to the equation system due to the presence of multiple local minima in the underlying cost function.
2. The application of M2 such that global best fit solutions are located (via optimisation) leads to significant over de-trending of TI distributions to a degree which implies *this method should no longer be used for this purpose.*

Put simply, the equation system for M2 results in over-fitting. One might be tempted to argue that the iterative approach to M2 can still be used, despite its flaws, since it has been found to produce reasonable results in practice. However, such an approach is clearly contradictory since it would rely on knowingly using sub-optimal solutions to the equation system. In addition, the cost function itself will vary between sites, and hence, in some cases, the iterative method may converge consistently to the true global solution and so exhibit the same significant over de-trending tendencies of the optimisation based solutions. Therefore, with respect to both principle and practice, neither implementation of M2 can be considered suitable for de-trending site data. As such, M2 will no longer be used in the current paper.

3.3.3 | Discussion

Having determined that M2 is no longer a viable method for moment based de-trending, M1 necessarily becomes the most attractive existing approach. However, M1 has been shown to behave somewhat conservatively,[1] leaving open the possibility that other methods might be developed which better recreate the raw data de-trended distributions when fitting to moment values. In this context, there are two important considerations which should be kept at the fore-front of such discussions:

1. Having taken raw data and processed it into its first two statistical moments only, almost all of the underlying information has been lost. As such, it is not immediately clear how close one should expect to be able to come to recreating distributions generated using the full information present in raw data. Indeed, M1 may already be maximising the potential for de-trending on the considered moment data.
2. Along these same lines, and as touched upon in Section 1, fitting methods have associated with them given levels of smoothing which, in the case of splines, is related to the degree of constituent polynomials. The choice of degree is often linked to the need for a solvable system of equations, rather than the characteristics of the function being recreated. Therefore, it is relevant to ask whether the degree chosen in M1 is optimal for the given task and, for any newly developed method, what an appropriate level of smoothing is to best identify the available information in moment data, limited though it may be.

The following GP implementation is developed with these points in mind, in particular, with a view to providing a principled approach to determining the correct amount of smoothing to apply, and comparing results with those of M1.

4 | A GP REGRESSION APPROACH

GPs have been studied as a regression technique since the early 90s. Initially discovered as being a limiting case of some types of neural networks, as the number of hidden layers becomes infinite,[9] they have since been shown to be a flexible and robust regression technique and have been applied to many different tasks,[10] including in the wind energy context.[11,12]

GP regression is a probabilistic technique for fitting a function to measured data. The set of measured values is assumed to have been produced by a function drawn from a GP prior distribution, potentially with additive noise present. The measured data itself are generally used to determine the parameters which define the prior, and noise variance, in a maximum likelihood procedure. This gives a multivariate Gaussian distribution between noisy measurements and underlying function values over any set of input points. Conditioning on measured values results in a posterior GP, from which the mean values at prediction points are interpreted as giving a best fit to the data, with standard deviations providing confidence intervals about the best fit. The following section makes this procedure explicit.

4.1 | GP regression equations

Let measured data occurring at time, t, be denoted y_t. It is assumed that the y_t are composed of the true underlying function values, f_t, corrupted by zero mean additive and independent Gaussian noise, ω_t. Therefore we have

$$y_t = f_t + \omega_t, \tag{30}$$

with $\omega_t \sim N(0, \sigma_\omega^2)$. The function values, f_t, are assumed to have been drawn from a zero mean GP prior. Thus, in order to specify the prior covariance of the f_t, it is necessary to specify a covariance structure, k, such that

$$k(f_{t_1}, f_{t_2}) = \mathbb{E}(f_{t_1} f_{t_2}). \tag{31}$$

It is standard practice to model the covariance of function outputs as being a function of their input values, $k(f_{t_1}, f_{t_2}) = k(t_1, t_2)$. This makes intuitive sense for continuous functions. The most common model for k is the squared-exponential covariance function,

$$k(t_1, t_2) = a \exp\left(-\frac{1}{2l}(t_1 - t_2)^2\right), \tag{32}$$

where a and l are hyperparameters which determine the amplitude and lengthscale (smoothness) of the covariance function, respectively. It is this covariance function on which the current work is based. As well as a and l, the noise variance σ^2 is also represented by a hyperparameter, ξ. Thus, while the general structure of the covariance model is specified, the parameters which populate it are generally learned from measured data.

Having specified a covariance function, covariance matrices can be formed for both measurement and prediction points. Covariance matrices are defined as follows for any two specified vectors,

$$[K_{\mathbf{ZZ}^*}]_{ij} = k\left([\mathbf{Z}]_i, [\mathbf{Z}^*]_j\right). \tag{33}$$

Note that all vectors are column vectors. Letting $\mathbf{Z}$ and $\mathbf{Z}^*$ be vectors of function input values at measurement and prediction points, respectively, the covariance matrices for prediction values and between measured and predicted values are then $K_{\mathbf{Z}^*\mathbf{Z}^*}$, $K_{\mathbf{Z}^*\mathbf{Z}}$ and $K_{\mathbf{ZZ}^*}$. The covariance matrix for measured values, denoted Q, has an additional contribution from the independent Gaussian noise term,

$$Q = K_{\mathbf{ZZ}} + \xi I. \tag{34}$$

Determining the GP prior is normally achieved by performing a maximum likelihood optimisation with respect to the measured data, Y_Z, in order to find the most likely values for hyperparameters a, l, and ξ given the observed data. In the current case however, it will be shown that hyperparameter values can be dealt with without needing to implement numerical procedures.

Having determined a GP prior, predictions in the posterior are obtained via the multivariate Gaussian conditional distribution. For a given vector of input points, $\mathbf{Z}^*$, the mean vector and covariance matrix of function outputs in the posterior are

$$\boldsymbol{\mu}_{\mathbf{Z}^*} = K_{\mathbf{Z}^*\mathbf{Z}}Q^{-1}\mathbf{Y}_{\mathbf{Z}} \tag{35}$$

and

$$\Sigma_{\mathbf{Z}^*} = K_{\mathbf{Z}^*\mathbf{Z}^*} - K_{\mathbf{Z}^*\mathbf{Z}}Q^{-1}K_{\mathbf{Z}\mathbf{Z}^*}, \tag{36}$$

respectively. As previously stated, $\boldsymbol{\mu}_{\mathbf{Z}^*}$ values are interpreted as being the best fit to the data at each point. Standard deviations, $\boldsymbol{\sigma}_{\mathbf{Z}^*}$, can be obtained as the square roots of the diagonal terms in $\Sigma_{\mathbf{Z}^*}$. Ninety-five percent confidence intervals are then given by $\boldsymbol{\mu}_{\mathbf{Z}^*} \pm 2\boldsymbol{\sigma}_{\mathbf{Z}^*}$. The above assumes that both f and Y_Z are zero mean; in practice, this requirement is achieved by subtracting the arithmetic mean of measured values from the data prior to regression, with the same value then added back on after predictions are made.

4.2 | GP moment de-trending

The current section considers the application of GP regression to wind site turbulence de-trending. Some work is required in order for GP regression to be applicable to the current problem. The most crucial task being to link the GP approach with the rigorous definition of a trend as given in Section 1.

It is proposed that GP theory, and its inherent flexibility, be leveraged to allow for the relationship between mean and raw data values to be properly accommodated within the model. This will allow a curve to be fitted, but in such a way that the composition of mean values and their relationship to individual function values is properly accounted for. The possibility for the approach developed here was inspired by previous work in GP theory which allows for regression models based on data not directly measuring the function being fitted. For example, it has been shown that function measurements can be used to developed a GP regression model for predicting the derivative of that function.[13] The key to doing this being to properly capture data-function relationships through a careful consideration of covariance structure. More generally, a proper consideration of covariance structure is known to play a key role in the effective application of these techniques.[14]

The same definitions and assumptions of Section 2.1, in particular the independence of g and s, are applied again here for the situation where Equation 30 takes the form $y_t = g_t + s_t$. First, the covariance properties of mean values, $M_{i\underline{T}}$, and individual trend function values, g_t, are considered. Since underlying function values in the GP are assumed to be zero mean, it follows that so are the $M_{i\underline{T}}$ (dealing with this in practice will be described below). In addition, the s contribution is assumed to take the form of noise. Observe that under these conditions,

$$\text{Var}\left(M_{i\underline{T}}, M_{j\underline{T}}\right) = \mathbb{E}\left(M_{i\underline{T}}M_{j\underline{T}}\right) \tag{37}$$

$$= \frac{1}{N^2}\sum_{t'\in i\underline{T}}\sum_{t^*\in j\underline{T}}k_g(t',t^*) + \delta_{ij}\frac{\xi}{N}. \tag{38}$$

The above expression suggests a form of covariance function between measured mean values. In this case, it has been found that the 'noise' contributions to covariance, ξ/N in the above expression, are in fact negligible. This follows from the fact that the left-hand expression in Equation 38 has been found to take values between 0.7 and 1 and that ξ values tend to lie between 0.5 and 1 for wind speed data. Given that wind speed measurements tend to be made at 1 or 2 Hz frequencies (or higher), N is commonly equal to 600 or 1200 (assuming 10-min capture windows, as is standard) and can be larger. Hence, the term ξ/N will generally be equal to 0.0017 or less, 2 orders of magnitude smaller than the dominant covariance term. This observation makes intuitive sense, since taking an average will result in the higher frequency, more noise like, signal contributions tending to zero. This allows for the latter term in Equation 38 to be ignored, reducing the number of hyperparameters required in this case. The following *mean values* covariance function can therefore be defined as

$$k_M(i,j) = \frac{1}{N^2}\sum_{t'\in i\underline{T}}\sum_{t^*\in j\underline{T}}k_g(t',t^*). \tag{39}$$

Similarly, and again using independence of g and s, one can derive the following covariance function between mean, $M_{i\underline{T}}$, and individual function values at a given point in time, g_t,

$$k_{gM}(t,i) = \frac{1}{N}\sum_{t'\in i\underline{T}}k_g(t,t'). \tag{40}$$

The predictive formula of Equation 35 can now be applied in the current case, with covariance matrices defined by the relevant covariance functions to transfer information from measured mean values to function predictions at individual timesteps. Explicitly, let

$$\mathbf{M} = \begin{bmatrix} M_{1\underline{T}} \\ \vdots \\ M_{m\underline{T}} \end{bmatrix}, \tag{41}$$

then for a given time window, $i\underline{T}$, the GP regression prediction for the vector of values taken by g in $i\underline{T}$ is

$$\boldsymbol{\mu}_{i\underline{T}}^{g} = K_{gM} K_M^{-1} \mathbf{M}, \tag{42}$$

where the matrix K_{gM} has entries given by $k_{gM}(t, i)$ for $t \in i\underline{T}$ and $i \in \{1, 2, \dots, m\}$ and K_M has entries given by $k_M(i, j)$ for $i, j \in \{1, 2, \dots, m\}$. While the noise hyperparameter has been removed, the amplitude and lengthscale hyperparameters are still present in the above expression. The removal of the noise term, in fact, also allows for cancellation of the amplitude hyperparameter, since

$$K_{gM} = a\hat{K}_{gM}, \tag{43}$$

$$K_M = a\hat{K}_M, \tag{44}$$

with the $\hat{K}$ matrices obtained by setting the amplitude equal to 1 in their constituent covariance terms. Re-examining Equation 42, it can be seen that the amplitude values cancel, removing any dependency on a in the current case. Note that this is only possible here due to the averaging process having removed the noise contribution normally present during GP regression. The GP predictions in this case are therefore obtained using the following expression:

$$\boldsymbol{\mu}_{i\underline{T}}^{g} = \hat{K}_{gM} \hat{K}_M^{-1} \mathbf{M}. \tag{45}$$

This formulation again assumes a zero mean GP, and so, as before, in practice, the arithmetic mean of the data (the values in $\mathbf{M}$ for the current case) is first subtracted from $\mathbf{M}$ before the regression equation is applied—with this same value then added on again afterwards. Given the removal of a and ξ hyperparameters, only the lengthscale l remains to be determined. It is shown in the following section how this can be done.

4.2.1 | Determining the lengthscale hyperparameter

The squared-exponential covariance function, Equation 32, has the lengthscale hyperparameter, l, which determines how fast covariances decay between output values as the input values move away from each other. Intuitively, one can see that large values of l correspond to functions which vary slowly across the domain, and small values to those which vary quickly. The value of l therefore dictates the amount of smoothing present when fitting predictions to measured data. For trend fitting, it is therefore necessary to set l such that it corresponds to a level of smoothing appropriate to capture the Fourier modes present in the signal with period greater than or equal to the size of the capture window, T; this being the definition of a trend in the current work.

The key to linking lengthscales and the trend definition turns out to be the more rigorous definition of lengthscale given by Adler,[10,15]

$$\mathbb{E}[N_u] = \frac{1}{2\pi} \sqrt{\frac{-k''(0)}{k(0)}} \exp\left(-\frac{u^2}{2k(0)}\right). \tag{46}$$

N_u is the number of upcrossings of a level u on the unit interval, and k is the covariance function of the considered (zero mean) GP. l can be isolated by inserting Equation 32 into Equation 46, setting $u = 0$ and rearranging to obtain

$$l = \left(\frac{1}{2\pi \mathbb{E}[N_0]}\right)^2. \tag{47}$$
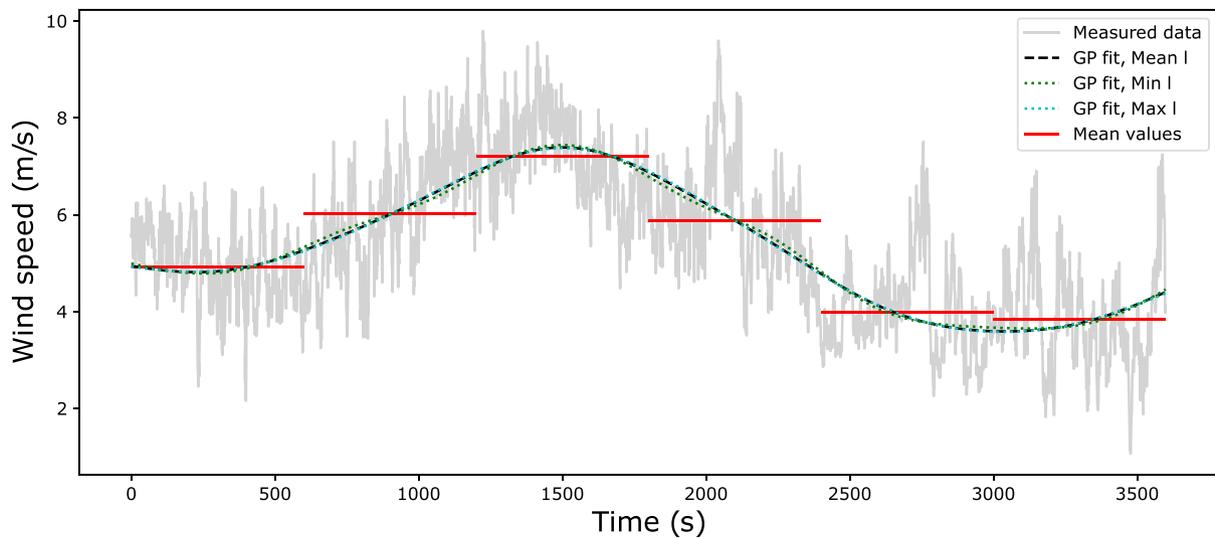
FIGURE 7 GP trend predictions from moment regression on mean values using the min, max and mean lengthscale hyperparameter values learned from wind site data. Underlying raw data are also shown for comparison [Colour figure can be viewed at wileyonlinelibrary.com]

In practice ,1-h duration filtered trends formed the basis for counting numbers of up-crossings, this being the longest duration over which some level of stationarity in the wind field might be assumed to hold (see Section 3). For each 1-hr window, the mean value was subtracted and up-crossings of zero counted. This was continued throughout a given time series and the mean number over an hour, $\mathbb{E}[N_{(0,1hr)}]$, computed. From this value, the lengthscale is determined as

$$l = \left(\frac{3600}{2\pi\mathbb{E}[N_{(0,1hr)}]} \right)^2, \tag{48}$$

which can be seen to be consistent with Equation 47 when one takes the unit interval stipulation of the definition into account. Note that the numerator in Equation 48 is expressed in seconds to be consistent with the units of GP input variables (Equation 32). Hence, lengthscale values can be systematically calculated across each of the sites for which raw data are available. This was therefore undertaken, with resulting values of l lying between 3.8×10^4 and 9.7×10^4, with an average value of 7.7×10^4. From experience, lengthscale hyperparameters generally need to see variations by an order of magnitude or more before any significant changes are discernible during regression. This was found to hold here also, as shown in Figure 7 where GP predictions are plotted when using the moment de-trending methodology with min, max and mean lengthscale values. Differences between predicted trends are minimal, with this same result found to hold in general, and hence, it can be concluded that the GP fits are not highly sensitive to lengthscale values. The implication being that the mean value of 7.7×10^4 can be taken as a suitable lengthscale value to encode the smoothness properties of real trend components (from where this value was derived), while also being applicable across the range of sites. This generality is important since it implies the GP moment methodology can be applied at sites for which raw data are not available by using this mean l value learned from existing raw data.

Thus, the prediction equation in the moment case (Equation 45) has been reduced to a deterministic linear transformation of measured mean values, with no requirement for optimisation-based parameter setting and, crucially, with a quantity of smoothing known to be consistent with wind site data trend components.

5 | RESULTS AND COMPARISONS

The GP and M1 moment de-trending methodologies were both run across all available sites listed in Table 1; six bins were fitted to simultaneously in each case. Figure 8 shows a typical example of trends predicted by these two methods. It can be seen that the GP and M1 methods result in almost identical fits across the majority of bins, with significant differences only apparent in a single bin. This was found to hold in general and appears to be caused by the different ways these methods behave at endpoints. With respect to de-trending, these differences turn out to only have a minor effect, as shown in the example TI distribution for Hanford and Sletringen shown in Figure 9.[#]

[#]In this case, the M1 distributions are those resulting from removing the full trend variance from measured values, rather than just the linear component. However, the differences between the linear and full variance approaches to M1 are almost indistinguishable in the resulting distributions. This is clear when comparing the M1 results in Figures 3 and 9.
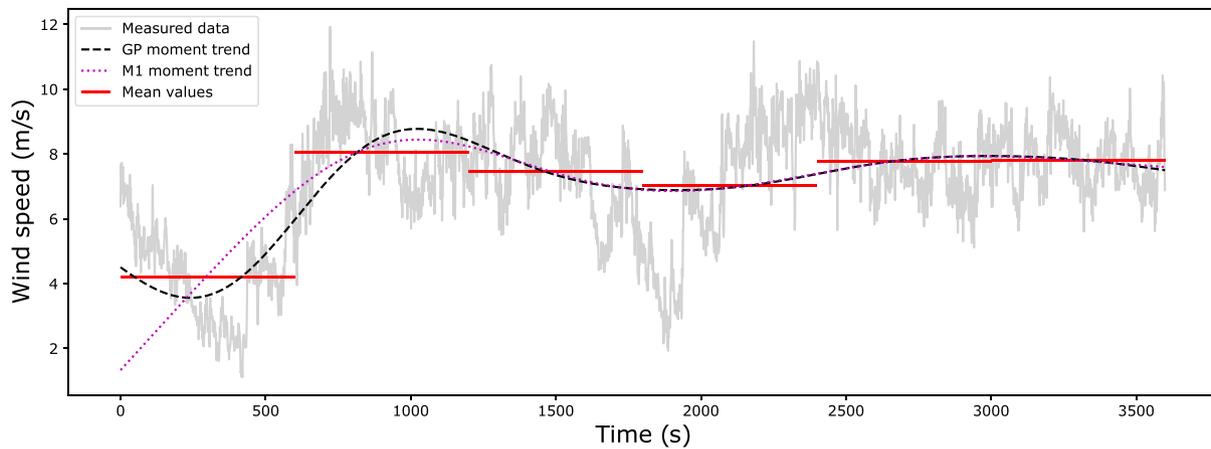
FIGURE 8 Time series regression examples for both M1 and GP moment de-trending across 1 h of mean data values. Underlying raw data are also shown for comparison [Colour figure can be viewed at wileyonlinelibrary.com]
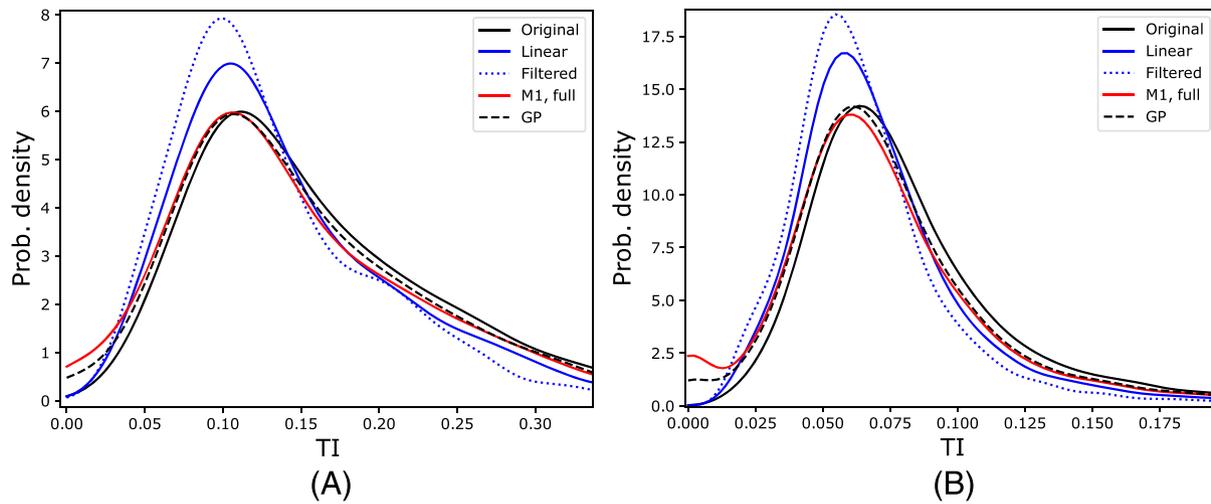
FIGURE 9 Moment de-trending using the GP method and Model 1 for (A) Hanford and (B) Sletringen datasets. For comparison, the original distributions and those obtained from linear- and filter-based de-trending using raw data are also shown [Colour figure can be viewed at wileyonlinelibrary.com]

It can be seen that the GP and M1 de-trended distributions are very similar, with M1 managing some small amount of de-trending above that managed by the GP. However, M1 in these examples also displays an increased tendency to over de-trend as compared to the GP. The same pattern is seen in general when considering the ODT and EMD measures across all sites, as shown in Figure 10. In all cases, the ODT measure for M1 is larger than that of the GP but never increases much beyond 3% and hence does not represent a significant issue for the viability of either method. With respect to EMD, it can be seen explicitly that the difference between linear or full variance de-trending with M1 is almost negligible. The observation that M1 carries out some small amount of de-trending beyond that of the GP is also clear, but, again, the differences are slight. Both the GP and M1 approaches can be seen to have an EMD less than that of linear de-trending on raw data, which is in turn less than that of filtered de-trending on raw data. Hence, both methods have been found to be conservative de-trending approaches overall which result in very similar predicted distributions.

In both cases, the over de-trending present can be seen to occur towards the lowest values of TI. This issue can therefore be dealt with by acknowledging this and simply applying a cutoff for post moment de-trended TI values within which the adjusted TIs are returned to their pre de-trending values. Based on the ODT results here, it is recommended that all de-trended TI values falling within the first 5% (conservatively) of probability mass for the original TI distribution are returned to their un-detrended values in order to minimise cases of over de-trending for both M1 and GP implementations.

Assuming these over de-trending cases are handled as described above, there is little difference remaining between the M1 and GP moment de-trending methods. Returning to the points raised in Section 3.3.3, the current results indicate that M1 is in fact behaving almost identically to
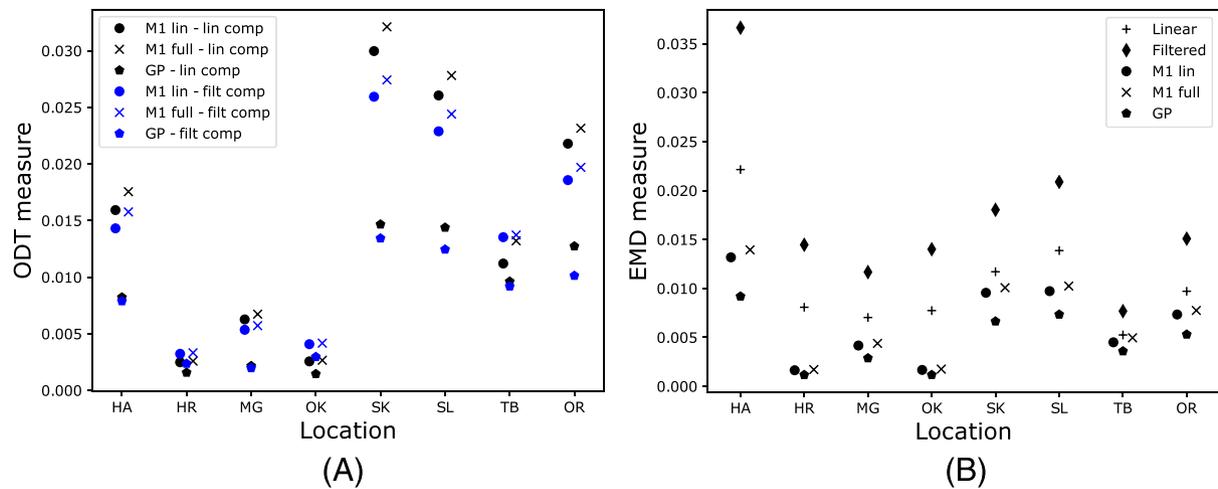
FIGURE 10 Performance results in the form of over de-trending (ODT) and earth-movers distance (EMD) measures for Model 1 and GP moment de-trending in subplots (A) and (B), respectively [Colour figure can be viewed at wileyonlinelibrary.com]

a method which specifically takes account of information relating to the appropriate level of smoothing to apply when fitting to moment data in order to identify trend components in wind site data. This can be interpreted as a validation of M1 overall and, in particular, of the order of polynomials fitted within each bin. Additionally, it indicates that M1 may in fact be doing as well as is possible given the information being fitted to in the current formulation of this problem, with the further implication that improved de-trending performance in the future may require leveraging of additional information, as opposed to simply creating refined techniques which use this same information. The GP moment de-trending methodology developed here has served the important function of allowing for characteristics of trend contributions in wind site data to be taken into account when predicting variance contributions and, in doing so, has allowed for the possible levels of de-trending obtainable from moment data, within the current problem formulation, to be considered. However, given this method has significant theoretical overheads as compared to M1, coupled to the fact that its performance in terms of de-trending is almost the same as M1, leads to the conclusion that M1 should be recommended as the best method for practical de-trending of wind site data at this time, having been validated by the GP approach.

6 | CONCLUSIONS

This paper has considered the problem of retrospective de-trending of wind site TI distributions using moment data only. Existing approaches, M1 and M2,[1] were applied and analysed in detail. This analysis showed that M1 behaves as described in the literature but that M2 has fundamental issues associated with it that make it necessary for this method to be abandoned as an approach to wind site turbulence de-trending. A GP regression approach to moment de-trending was then developed, which captures the smoothness associated with trend components in wind data via a lengthscale hyperparameter within the covariance structure. From testing and comparisons across a range of sites, it was shown that the M1 and GP methods result in very similar de-trended distributions, with only small instances of over de-trending towards low TI values. It was argued that these cases can be avoided by returning all TIs within a given range of probability mass to their un-detrended values. Overall, the similarity in performance between these two methods was interpreted as a validation of M1 and the quantity of smoothing it applies when fitting. Given the theoretical overheads of the GP implementation, it is concluded that M1 should be recommended as the method of choice for moment-based turbulence de-trending at this time.

PEER REVIEW
The peer review history for this article is available at https://publons.com/publon/10.1002/we.2614.

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

REFERENCES

1. 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.
2. Ziegler L, Gonzalez E, Rubert T, Smolka JUM. Lifetime extension of onshore wind turbines: A review covering Germany, Spain, Denmark, and the UK. *Renew Sustain Energy Rev*. 2018;82:1261-1271.
3. Bouty C, Schafhirt S, Ziegler S, Muskulus M. Lifetime extension for large offshore wind farms: Is it enough to reassess fatigue for selected design positions? *Energy Proc*. 2017;137:523-530.
4. Chevalier G. Python code. Filtering signal with a butterworth low-pass filter and plotting the STFT with a Hanning window and Laplace transform. https://github.com/guillaume-chevalier/filtering-stft-and-laplace-transform. Last accessed: 16-06-2020.
5. IEC 61400-1:2019. Wind energy generation systems—Part 1: Design requirements. *Standard*, Geneva, Switzerland, International Electrotechnical Commission; 2019.
6. Wasserstein distance, python documentation. https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.wasserstein_distance.html, Last accessed: 1-06-2020.
7. Ramdas A, Trillos NG, Cuturi M. On Wasserstein two sample testing and related families of nonparametric tests. *Entropy*. 2017;19(47):1-15.
8. Escalante Soberanis MA, Merida W. Regarding the influence of the Van der Hoven spectrum on wind energy applications in the meteorological meso-scale and microscale. *Renew Energy*. 2015;81:286-292.
9. Neal RM. Priors for infinite networks. Technical Report CRG-TR-94-1, Toronto, University of Toronto; 1994.
10. Rasmussen CE, Williams CKI. *Gaussian Processes for Machine Learning*. Cambridge, MA, USA: MIT Press; 2006.
11. Stock-Williams C, Mazoyer P, Combrexelle S. Wind field reconstruction from lidar measurements at high frequency using machine learning. *J Phys Conf. Ser*. 2018;1102:12033.
12. Rogers TJ, Gardner P, Dervilis N, et al. Probabilistic modelling of wind turbine power curves with application of heteroscedastic Gaussian process regression. *Renew Energy*. 2020;148:1124-1136.
13. Solak E, Murray-Smith R, Leithead WE, Leith DJ, Rasmussen CE. Derivative observations in gaussian process models of dynamic systems. *Adv Neural Info Process Syst*. 2003;15:1057-1064.
14. Wilson AG, Adams RP. Gaussian process kernels for pattern discovery and extrapolation. In *Proceedings of the 30th International Conference on Machine Learning, JMLR*. Atlanta, Georgia, USA: W&CP; 2013:28.
15. Adler RJ. *The Geometry of Random Fields*. Chichester: Wiley; 1981.