Predicting Underwater Noise Spectra Dominated by Wind Turbine Contributions

Andrea Trucco, *Senior Member, IEEE*

***Abstract*—The study of the impact on the marine ecosystem of an offshore wind farm benefits from the knowledge of the underwater noise observed at a single turbine, as the wind speed varies. The calculation of the noise spectral average at a given wind speed requires many recordings, each acquired in a limited time interval: an extremely time-consuming process. This study investigated how to approach the spectral average using only very few noise recordings for each wind speed, leveraging supervised and unsupervised machine learning techniques. Three different prediction methods, based on mean and interpolation, principal component analysis (PCA), and nonnegative matrix factorization, in combination with four techniques for coefficient estimation as the wind varies, are tested. Prediction based on principal component analysis, combined with Gaussian process regression, outperforms other methods in all three case studies considered. The latter, in addition to the problem described above, include the prediction of the noise spectrum: at wind speeds where no noise recordings are available, and using a few recordings acquired at another (nominally identical) wind turbine.**

***Index Terms*—Offshore wind turbine, principal component analysis (PCA), spectral prediction, supervised and unsupervised learning, underwater noise.**

I. Introduction

As OFFSHORE wind farms (OWFs) become more widespread, the impact on the ecosystem of underwater noise they generate is a cause for concern [1], [2], [3]. In addition to the installation and decommissioning of a plant, which can produce high noise levels for limited periods of time [1], [4], the noise generated during the normal operation of the OWF (i.e., the operational noise) can affect the marine ecosystem for a very long time and is, therefore, particularly worrying [1]. In recent years, the operational noise of an OWF has been the subject of several investigations [2], [5], [6], [7], showing that, within a given area, the turbines' contribution to the underwater noise spectrum can be dominant, especially as the wind speed increases. To assess the impact of an OWF during its operation, it is necessary to know the power spectral density (PSD) of noise at different points in a given marine area [1], [4], [5], [7] as a function of wind speed. To achieve such knowledge, it is useful to characterize each turbine in the plant acoustically [2], [5], [8], [9]. The noise model near a wind turbine can also serve as a basis for estimating the overall noise of a hypothetical OWF at an early design stage [2], [4], [7].

This article focuses on the measurement and prediction of underwater noise spectra observed near a wind turbine, including ambient noise and turbine contributions. Previous measurements at sea have shown that the turbine's role [1], [8], [9] is mainly to insert some spectral lines, often below 1 kHz, into the typical broadband spectrum of ambient noise. Frequencies and amplitudes of such tonal components exhibit a complex relationship with wind speed, especially for turbines operating at variable rates of revolution [8], [9].

Let us assume for a while that a large number of noise spectra measured near the turbine, for each wind speed (having suitably discretized this variable), is available. Given the stochastic nature of noise, the average of all the spectra (i.e., the spectral average) measured at a given wind speed represents the reference option to express the underwater noise generated by the turbine and natural sources at the speed considered. The question this article contributes to answering is: How to predict such a spectral average if only very few noise spectra are available at the considered speed or if no spectra are available at that speed and only very few spectra are available at different speeds? The rationale behind the question concerns the time necessary to acquire many noise spectra for each wind speed (some speeds being rather infrequent [10]) and the costs associated with this operation. An additional question addressed is: can the few noise spectra acquired near a given turbine effectively predict the spectral average of the noise near another OWF turbine, provided that the turbines are two specimens of the same model?

Although the literature concerning the measurement and characterization of underwater noise generated by a wind turbine or an entire OWF is growing rapidly,[1] to the best of the author's knowledge, no one has yet investigated how best to make use of few noise spectra acquired near a turbine, what is reasonable to expect from them, and how well they can predict the noise near another turbine. This topic is important because the spectral prediction techniques studied in this work allow for approximate (but reasonably accurate) knowledge of spectral averages, using much less time and cost than would be required to acquire a data set containing many noise spectra for each wind speed.

Manuscript received 13 September 2023; revised 4 June 2024; accepted 4 June 2024. This work was supported by the Joint Programming Initiative Healthy and Productive Seas and Oceans (a pan-European intergovernmental platform) through the Project PURE WIND—Impact of sound on marine ecosystems from offshore wind energy generation.

Associate Editor: V. Piscopo.

The author is with the Department of Electrical, Electronic, Telecommunications Engineering, and Naval Architecture, University of Genoa, 16145 Genoa, Italy (e-mail: trucco@ieee.org).

Digital Object Identifier 10.1109/JOE.2024.3415753

[1]The authors of [2], [3], and [5] present recent reviews on this topic. Works [6] and [7] are examples of recent contributions.

To compare different spectral prediction strategies, this article uses the data set found in the supplementary material of [8]. It contains many spectra acquired near a given turbine, each spectrum obtained by processing the underwater noise recorded over a 10-min period. With few exceptions, for every wind speed, discretized at a step of 1 m/s, the data set contains more than 100 spectra. They are used here as follows: a few spectra for each wind speed are randomly selected and set aside; the average among the (many) remaining spectra for each speed represents the spectral average, i.e., the target that the prediction strategies considered try to approximate using the few spectra set aside, using all or part of them. Moreover, the data set contains the recordings described previously for two nominally identical OWF turbines [8], also allowing comparison of the noise acquired near each.

The simplest prediction strategy is to average the few available spectra for each wind speed. If no spectra are available at certain speeds, a linear interpolation of the results obtained at the remaining speeds can be performed. However, there is no guarantee that this is the best strategy. A possible alternative is a two-stage procedure: 1) to analyze the few noise spectra available to extract the salient patterns (or basis spectra) they have in common; and 2) to synthesize a prediction (trying to approximate the spectral average at a given speed) by combining such basis spectra appropriately. For the analysis stage, two unsupervised learning techniques were considered, which are widely used in exploratory data analysis to achieve dimensionality reduction [11]: principal component analysis (PCA) and nonnegative matrix factorization (NMF). In both, a noise spectrum is modeled as a linear combination of the basis spectra, the number of which should be suitably set. Obviously, the coefficients weighting the basis spectra must depend on the wind speed: the value of a given coefficient as a function of speed can be determined by a fitting or regression technique, fed by the values the coefficient assumed in the analysis stage. This task will be referred to as coefficient estimation [11], [12]. To test various possibilities, curve fitting using second-order polynomials [12] and smoothing splines [13] was considered, as well as supervised learning regression [11], [14] based on random forests and Gaussian processes.

This study shows that an accurate design of the prediction pipeline produces results significantly closer to the target, on average, and with less variance, than those produced by averaging and interpolation (AAI). With reference to the mentioned data set, it is anticipated that by using five spectral noise samples for each wind speed, the average of the mean absolute log-spectral distance can be reduced to below 1.4 dB[2] and its standard deviation (std. dev.) to below 0.4 dB.

The rest of this article is organized as follows. Section II introduces the problem addressed, prediction methods, estimation techniques, and performance metrics. Section III presents the adopted data set and proposes a preparatory analysis. Section IV describes and discusses the results, including turbine data discrepancies, comparison of prediction strategies and estimation techniques, and assessment of some examples. Finally, Section V concludes this article.

II. Methods

A. Problem Statement

The general problem addressed in this article is how to predict the noise spectral average at a given wind speed, using very few spectral noise samples acquired at the same speed or, failing that, at speeds close to that of interest. Specifically, five spectral samples will be considered in the following.[3] Taking into account the characteristics of the available data set [8], the general problem is declined in three specific case studies, illustrated graphically in Fig. 1.

1) *Case study C1:* For each wind speed, the noise spectra acquired near a given turbine are allocated as follows. Five spectra chosen at random contribute to compose the so-called *available set* and will be used for prediction, while all the rest are used to calculate the spectral average [see Fig. 1(a)]. At each speed, the prediction will be compared with the spectral average to assess the quality of the result.
2) *Case study C2:* Same as case C1, except that for a subset of wind speeds, no noise spectra are used to compose the *available set* and compute the predictions. For each speed belonging to that subset, all spectra are used to compute the noise spectral average [see Fig. 1(b)].
3) *Case study C3:* For each wind speed, five spectra randomly chosen from those acquired near turbine Φ contribute to compose the *available set* and will be used for prediction. All noise spectra acquired near turbine Γ at the same wind speed are used to calculate the spectral average [see Fig. 1(c)]. The prediction computed with samples from turbine Φ will be compared with the spectral average from turbine Γ.

The quality of the result will be evaluated with the error metrics introduced in Section II-F. To understand which spectral prediction method provides the best results in the three mentioned case studies, the experiments must be repeated a great number of times, allowing the average and std. dev. of each error to be calculated.

B. Spectral Prediction by AAI

Let $\mathbf{x}_{j,i}$ be a noise spectrum among those composing the *available set* (i.e., those available to compute spectral predictions), where the index j, $j \in [1, m]$ refers to the wind speed w_j at which the spectrum was acquired, and the index i, $i \in [1, n]$ scans the n spectra available[4] for each speed w_j. Thus, the total number of noise spectra that compose the *available set* is $N = mn$. Each spectrum $\mathbf{x}_{j,i}$ is a row vector with p entries, which express the noise PSD, in linear scale, at p frequency values.

[2] This value increases to about 2.4 dB when the spectral averages of one turbine are predicted using five noise spectra, for each wind speed, recorded close to the other turbine.

[3] The impact on the results of having a different number of spectral samples is addressed in Section IV-D, where said number is a random variable for each wind speed.

[4] As anticipated, it is intended that n be a few units.

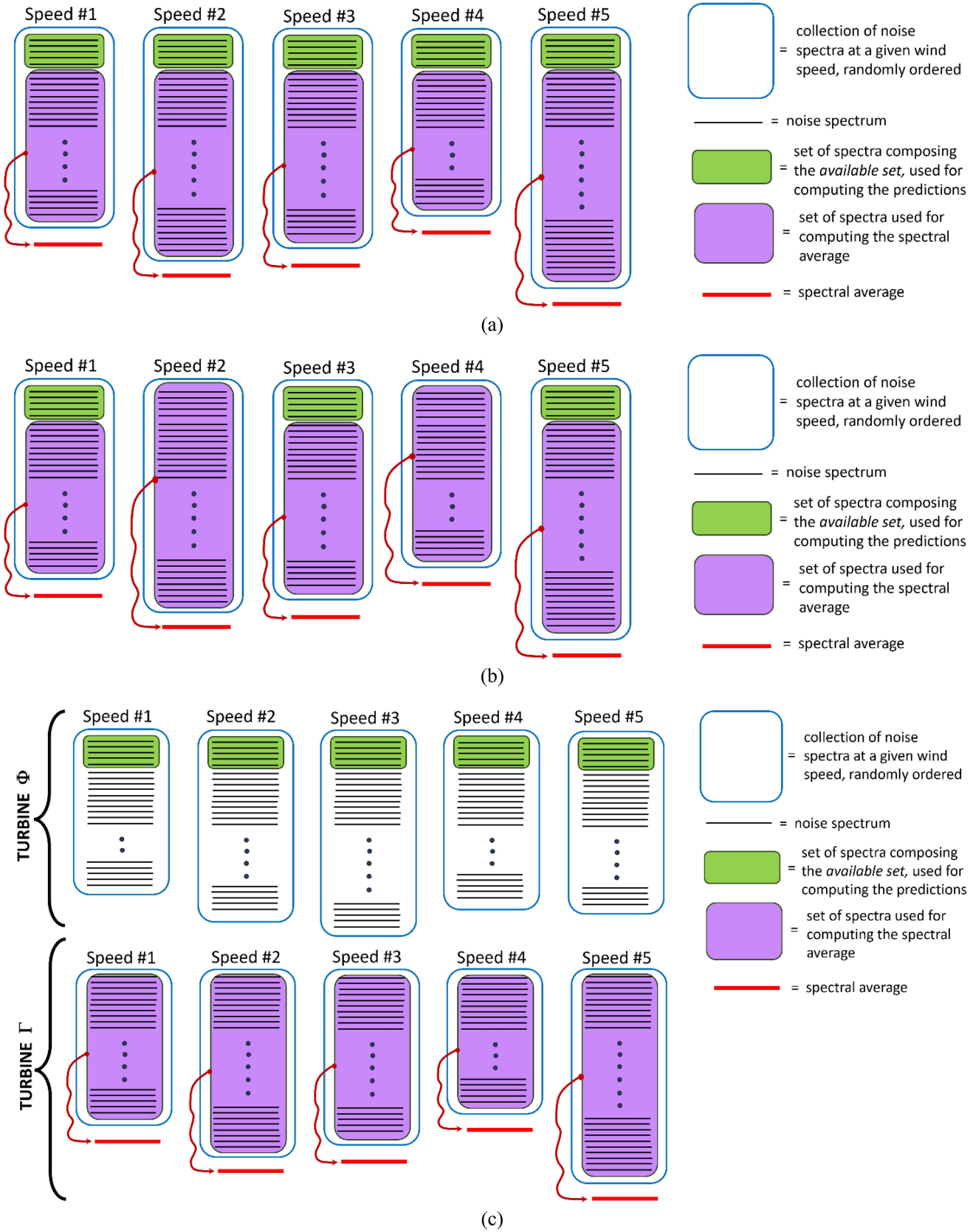


Fig. 1. Prediction of noise spectral averages using a few spectral samples. (a) Case study C1. (b) Case study C2. (c) Case study C3.

As shown in Fig. 1, the number of wind speeds considered in the *available set*, m, includes: all the speeds at which the noise spectra were acquired in the C1 and C3 cases; and a subset of them in the C2 case. Consequently, N will be greater in C1 and C3 than in C2. The simplest way to provide the spectral prediction at wind speed w, $\hat{\mathbf{s}}(w)$, is to compute the average, $\bar{\mathbf{x}}_j$, of the available spectra

$$\bar{\mathbf{x}}_j = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_{j,i} \tag{1}$$

for all the wind speeds w_j, $j \in [1, m]$, and use them as follows:

$$\hat{\mathbf{s}}(w) = \begin{cases} \bar{\mathbf{x}}_j, & \text{if } w = w_j \\ \mathcal{L}\left\{\bar{\mathbf{x}}_{j-1}, \bar{\mathbf{x}}_j\right\}, & \text{otherwise} \end{cases} \tag{2}$$

$\mathcal{L}\{\bar{\mathbf{x}}_{j-1}, \bar{\mathbf{x}}_j\}$ is an operator that, considering the distances separating w from w_{j-1} and w_j, performs the linear interpolation between $\bar{\mathbf{x}}_{j-1}$ and $\bar{\mathbf{x}}_j$, having chosen j such that $w_{j-1} < w < w_j$.

C. Spectral Prediction by PCA

Let $\mathbf{X}$ be the $N \times p$ matrix that contains all the N spectra of the *available set*, built by stacking a vector $\mathbf{x}_{j,i}$ for each row. Correspondingly, let $\mathbf{w}$ be the column vector the N entries of which are the wind speeds w_j related to the spectra $\mathbf{x}_{j,i}$ stacked in $\mathbf{X}$. Since PCA analysis [11] requires that the p variables represented in the data matrix $\mathbf{X}$ have zero mean, let us define the location vector $\boldsymbol{\mu}$ as a row vector obtained by calculating the sample mean along each column of $\mathbf{X}$. Then, by subtracting $\boldsymbol{\mu}$ from each row of $\mathbf{X}$, one obtains the matrix $\mathbf{X}_0$, the p variables of which have zero mean.

PCA uses the singular value decomposition to write $\mathbf{X}_0$ as

$$\mathbf{X}_0 = \mathbf{U}\mathbf{D}\mathbf{V}^T \tag{3}$$

where the superscript *T* indicates the transpose of a matrix, $\mathbf{U}$ is an $N \times p$ orthogonal matrix, $\mathbf{V}$ is a $p \times p$ orthogonal matrix, the columns of which are called the right singular vectors, and $\mathbf{D}$ is a $p \times p$ diagonal matrix, with diagonal elements

$$d_1 \geq d_2 \geq \ldots \geq d_p \geq 0$$

known as the singular values. The columns of the $N \times p$ matrix $\mathbf{U}\mathbf{D}$ are called principal components of $\mathbf{X}_0$. The sorting of the singular values ensures that the columns of $\mathbf{V}$ are not only orthogonal unit vectors representing the principal component directions but are also sorted in order of descending importance in explaining the data in $\mathbf{X}_0$.

On the basis of the above, it is possible to choose suitably a value q, $q < p$, which allows one to approximate the data matrix by keeping only the first q principal components and their associated directions

$$\mathbf{X}_0 \cong \boldsymbol{\Lambda}\boldsymbol{V}_q^T \tag{4}$$

where $\boldsymbol{\Lambda}$ is an $N \times q$ matrix that contains only the first q columns of the product $\mathbf{U}\mathbf{D}$ and $\mathbf{V}_q$ is a $p \times q$ matrix that contains only the first q columns of $\mathbf{V}$. The columns of $\mathbf{V}_q$ represent a set of q basis spectra that, linearly combined through the weight coefficients contained in each row of $\boldsymbol{\Lambda}$, approximate the spectra with zero-mean variables stacked in $\mathbf{X}_0$.[5] A given original spectrum can be approximated as

$$\mathbf{x}_{j,i} \cong \boldsymbol{\mu} + \boldsymbol{\lambda}_\eta \mathbf{V}_q^T \tag{5}$$

where $\boldsymbol{\lambda}_\eta$ is the ηth row of the matrix $\boldsymbol{\Lambda}$, provided that $\mathbf{x}_{j,i}$ is the ηth row of $\mathbf{X}$.

The basic idea in bounding the number of basis spectra to q is to keep the salient patterns that PCA identifies in the *available set* spectra, while eliminating patterns that represent specific features of such spectra, lacking generality. The noise spectral prediction at wind speed w, $\hat{\mathbf{s}}(w)$, can be computed by combining the q basis spectra through coefficients that must be properly estimated

$$\hat{\mathbf{s}}(w) = \left|\boldsymbol{\mu} + \hat{\boldsymbol{\lambda}}_w \mathbf{V}_q^T\right| \tag{6}$$

where $\hat{\boldsymbol{\lambda}}_w$ is a row vector of size q that contains such coefficients and $|\cdot|$ indicates the entrywise absolute value.[6] Each entry of $\hat{\boldsymbol{\lambda}}_w$ can be estimated on the basis of the values contained in the matrix $\boldsymbol{\Lambda}$. Specifically, the θth column of the matrix $\boldsymbol{\Lambda}$ contains the N values assumed by the θth coefficient, $\theta \in [1, q]$, in relation with the N samples of the *available set*. Since n samples are included for each w_j, $j \in [1, m]$, the θth column of $\boldsymbol{\Lambda}$ contains the n values the coefficient assumed for each wind speed. Recalling that the column vector $\mathbf{w}$ of size N contains the wind speeds at which the *available set* samples were acquired, the θth coefficient, $\hat{\boldsymbol{\lambda}}_w[\theta]$, can be estimated as follows:

$$\hat{\boldsymbol{\lambda}}_w\ [\theta] = \mathcal{R}\{\ w;\ \mathbf{w}, \boldsymbol{\Lambda}\,[:, \theta]\ \} \tag{7}$$

where $\boldsymbol{\Lambda}[:, \theta]$ is the θth column of $\boldsymbol{\Lambda}$ and $\mathcal{R}\{\cdot\}$ is an operator that, knowing the values $\boldsymbol{\Lambda}[:, \theta]$ the dependent variable assumes when the independent variable is $\mathbf{w}$, estimates the dependent variable when the independent variable is w. Such an operator may be a curve-fitting or a regression technique, as will be discussed in Section II-E.

D. Spectral Prediction by NMF

NMF is a technique to approximately factorize a data matrix $\mathbf{X}$, the entries of which are all nonnegative, into two matrices $\mathbf{W}$ and $\mathbf{H}$, also devoid of negative entries, achieving dimensionality reduction [11]. Since the PSDs contained in $\mathbf{X}$ ensure that this matrix is nonnegative, the NMF is suited to find a set of basis spectra that reveal the latent structure in the data, allowing noise spectra to be represented as the sum of nonnegative patterns. The conceptual differences between NMF and PCA were discussed in [15].

The NMF is commonly written as

$$\mathbf{X}^T \cong \mathbf{W}\mathbf{H} \tag{8}$$

where $\mathbf{W}$ is a $p \times r$ matrix, $\mathbf{H}$ is an $r \times N$ matrix, and $r < \min(p, N)$. In analogy with (4), the factorization can be

[5] The values in the basis spectra and the coefficients that weight them can be greater than, equal to, or less than zero. On the other hand, the product between the two matrices must generate zero-mean variables.

[6] Since it is not guaranteed that the predicted PSD is nonnegative at all frequencies, as suggested in [19], the absolute value of the results is considered.

rewritten as $\mathbf{X} \cong \mathbf{H}^T\mathbf{W}^T$: the columns of $\mathbf{W}$ represent a set of r basis (nonnegative) spectra that, summed through the weight coefficients contained in each column of $\mathbf{H}$, approximate the original spectra stacked in $\mathbf{X}$. Once r is set,[7] the factorization is performed by minimizing the quadratic distance between $\mathbf{X}^T$ and $\mathbf{WH}$, starting with random initial values for $\mathbf{W}$ and $\mathbf{H}$ and proceeding iteratively with the alternating least squares algorithm.

The noise spectral prediction at wind speed w, $\hat{\mathbf{s}}(w)$, can be obtained by combining the r basis spectra through coefficients that must be properly estimated

$$\hat{\mathbf{s}}(w) = \left|\hat{\mathbf{h}}_w \mathbf{W}^T\right| \tag{9}$$

where $\hat{\mathbf{h}}_w$ is a row vector of size r that contains such coefficients. Each of them can be estimated on the basis of the values contained in the matrix $\mathbf{H}^T$. Specifically, the ρth column of the matrix $\mathbf{H}^T$ contains the N values assumed by the ρth coefficient, $\rho \in [1, r]$, in relation to the N samples of the *available set*. Therefore, the ρth coefficient, $\hat{\mathbf{h}}_w[\rho]$, can be estimated as follows:

$$\hat{\mathbf{h}}_w[\rho] = \mathcal{R}\left\{\, w;\ \mathbf{w}, \mathbf{H}^T[:, \rho] \,\right\} \tag{10}$$

where $\mathbf{H}^T[:, \rho]$ is the ρth column of $\mathbf{H}^T$ and $\mathcal{R}\{\cdot\}$ is an operator that, knowing the values $\mathbf{H}^T[:, \rho]$ the dependent variable assumes when the independent variable is $\mathbf{w}$, estimates the dependent variable when the independent variable is w. Such an operator may be a curve-fitting or a regression technique, as will be discussed in Section II-E.

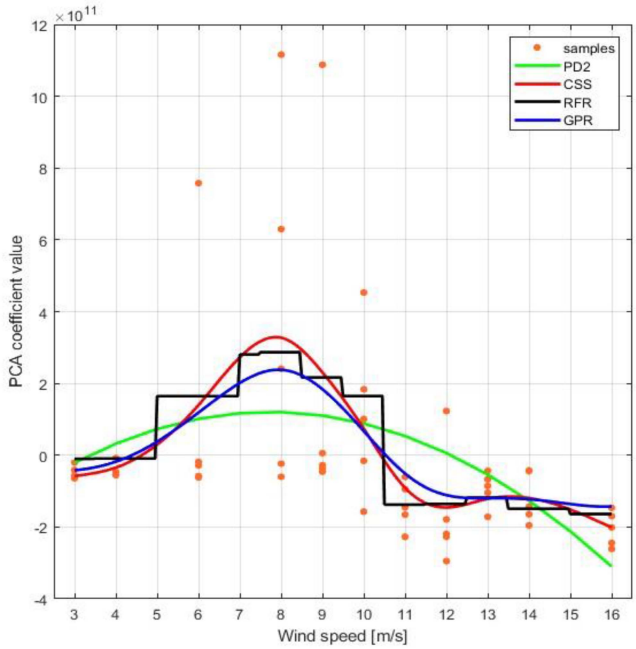


Fig. 2. Values assumed by the coefficient (orange dots) that weights a given PCA basis spectrum, as a function of the wind speed. The *available set* consisted of five noise spectra for each wind speed, from 3 to 16 m/s, in 1-m/s steps, excluding 5, 7, and 15 m/s, at which no spectra were included (i.e., an example of case study C2). The solid lines show the results obtained with the fitting and regression techniques described in the text.

E. Coefficient Estimation: Fitting and Regression Techniques

The analysis of the *available set*, either by PCA or NMF, yields a set of basis spectra that, combined linearly by weighting coefficients, approximate the original spectra. For each wind speed considered in the *available set*, the analysis generates n values of each coefficient. To achieve a noise spectral prediction depending on wind speed, for each coefficient, a function must be found that associates any speed value with only one coefficient value.

The problem, of which Fig. 2 shows an example, is the one formulated in (7) and (10) and can be approached from a mathematical or machine learning perspective. In the former case, curve-fitting techniques can be adopted; in the latter case, regression techniques derived from supervised learning can be used. Noting that the aim here is not to find the best technique for the problem at hand, but to test the sensitivity of the results against different techniques, both options are experimented, considering two techniques for each option. These techniques were chosen by simply trying to differentiate the categories to which they belong.

Among curve-fitting techniques, a polynomial with degree 2 (PD2) and a cubic smoothing spline (CSS) are considered. PD2 is a parametric model, the three parameters of which are set through a linear least-squares procedure [12]. Instead, CSS is a nonparametric model aimed at minimizing the sum of an error term and a roughness term [13]. The relative importance of the two terms is governed by a smooth factor that is tuned according the "de Boor Method" described in [16].

To avoid defining a priori a parametric form for the relationship between the independent and dependent variables, nonparametric supervised learning techniques are considered, among which random forest regression (RFR) and Gaussian process regression (GPR) are selected. RFR combines ensemble learning methods with decision trees to create multiple uncorrelated base estimators, the results of which will then be averaged to compute the final prediction [11]. In the problem at hand, the main hyperparameters are the number of trees in the ensemble, $\mathcal{E}$, and the minimum number of samples required to be at a leaf node, ℓ. GPR belongs to a class of probabilistic statistical models in which Gaussian processes are used to describe the uncertainty about a latent function [14], both as prior for Bayesian inference and as posterior distribution, obtained exploiting the data of the *available set* to update the prior. The choice of the prior's covariance function enables to encode assumptions on the smoothness and expected patterns of the latent function. Unlike previous techniques, where the minimization of a mean square error is performed, GPR maximizes a likelihood function to fit the model on the available data. Appendix A provides a more detailed description of these learning techniques as well as

[7] Unlike PCA, the patterns that added together approximate the spectra of the *available set* are not ordered by importance. Therefore, the choice of r must ensure a sufficient number of patterns to approximate the noise spectra with due accuracy and, at the same time, maintain generality. The choice of r will be discussed later.

additional information on the hyperparameters setting for RFR and the choice of the covariance function for GPR.

Fig. 2 shows an example of the estimates obtained with the four techniques described: while the curves generated by CSS, RFR, and GPR follow a common profile, the fitting with PD2 cannot go beyond the profile of a quadratic function. The RFR estimate has a stepwise profile because the values produced by the $\mathcal{E}$ trees (i.e., the values that are averaged to obtain the final estimate) vary discretely as wind speed changes.

F. Performance Assessment

Let $\mathbf{s}(w)$ be the noise spectral average at wind speed w, computed by using all the noise spectra not included in the *available set* (see the purple boxes and red lines in Fig. 1) and assuming that w is one of the speeds at which the noise spectra in the data set were acquired. Let $\hat{\mathbf{s}}(w)$ be the corresponding noise spectral prediction, obtained by one of the methods described in Sections II-B–II-D. Both the average and prediction are row vectors with p entries, which express the PSD (in linear scale) at p frequency values. Finally, let $\mathbf{s}_{\mathrm{dB}}(w)$ and $\hat{\mathbf{s}}_{\mathrm{dB}}(w)$ be the versions of $\mathbf{s}(w)$ and $\hat{\mathbf{s}}(w)$, respectively, in which the PSD is expressed in logarithmic scale, namely, in decibels.

To assess how close the prediction $\hat{\mathbf{s}}(w)$ is to the spectral average $\mathbf{s}(w)$, three error metrics will be considered [17], [18]: 1) the mean absolute log-spectral distance ($\mathrm{LD}_{\mathrm{ma}}$); 2) the root-mean-square log-spectral distance ($\mathrm{LD}_{\mathrm{rms}}$); and 3) the Itakura–Saito divergence (ISD). They can be computed as follows:

$$\mathrm{LD}_{\mathrm{ma}}(w) = \frac{1}{p}\sum_{b=1}^{p}\left|\hat{\mathbf{s}}_{\mathrm{dB}}(w)_b - \mathbf{s}_{\mathrm{dB}}(w)_b\right| \tag{11}$$

$$\mathrm{LD}_{\mathrm{rms}}(w) = \sqrt{\frac{1}{p}\sum_{b=1}^{p}\left(\hat{\mathbf{s}}_{\mathrm{dB}}(w)_b - \mathbf{s}_{\mathrm{dB}}(w)_b\right)^2} \tag{12}$$

$$\mathrm{ISD}(w) = \frac{1}{p}\sum_{b=1}^{p}\left(\frac{\mathbf{s}(w)_b}{\hat{\mathbf{s}}(w)_b} - \ln\frac{\mathbf{s}(w)_q}{\hat{\mathbf{s}}(w)_b} - 1\right) \tag{13}$$

where the subscript b, $b \in [1, p]$, indicates the bth entry of the vector .

Both $\mathrm{LD}_{\mathrm{ma}}$ and $\mathrm{LD}_{\mathrm{rms}}$ express average prediction errors in the same units as the spectra they process, i.e., decibels. However, while $\mathrm{LD}_{\mathrm{ma}}$ is just the sample mean of the distance between $\mathbf{s}_{\mathrm{dB}}(w)$ and $\hat{\mathbf{s}}_{\mathrm{dB}}(w)$ along frequency, $\mathrm{LD}_{\mathrm{rms}}$ assigns a relatively high weight to large errors and is more useful when large errors (e.g., mismatches in spectral lines) are particularly undesirable. ISD is a measure of the difference between the two spectra, equal to zero if the two spectra coincide, intended to reflect perceptual dissimilarity. In particular, ISD is known to provide more emphasis on spectral peaks than spectral valleys.

Another possibility for assessing spectra similarity is Pearson's correlation coefficient. Although it is scarcely considered in the acoustic field [17], it has wide application in other fields, such as biological spectrometry, chemical spectroscopy, and hyperspectral remote sensing. Appendix B elaborates on this, showing that the results obtained are aligned with those of the three metrics introduced earlier.

III. Data Set

The methods described in the previous section are tested using the data set provided as supplementary material in [8], where underwater noise near two identical turbines within the same OWF (Sheringham Shoal, about 20 km off the Norfolk coast in the U.K.) was collected for a period of 21 days, starting on April 18, 2013. The wind farm includes 88 Siemens SWT-3.6-107 turbines, supported by steel monopiles in a water depth of about 20 m, equipped with a three-bladed rotor that is controlled at variable rotational rate. The cut-in wind speed is about 3 m/s, the nominal power generation of 3.6 MW is reached at speed of 13 m/s, and the cutoff speed is 25 m/s. According to [8], the collection period allowed embracing of a full range of operational conditions.

The acoustic measurements were performed using two DSG-Ocean Acoustic Dataloggers, duly calibrated, deployed 50 m away from turbines A1 and A5 of the mentioned OWF[8] (see [8] for details). The underwater sound pressure acquired by each recorder was broken in 10-min periods, and Pangerc et al. [8] computed the related PSDs by Welch's method [19], adopting 1-s nonoverlapping segments and the Hanning window. For each 10-min period, a wind speed measurement, discretized in 1-m/s steps, was acquired 80 m above the sea surface, separately for the two turbines. The supplementary material of [8] contains all computed PSDs, from 0 to 3508 Hz at 1-Hz step, expressed in dB re 1μPa^2/Hz, sorted by simultaneous wind speed, separately for A1 and A5 turbines.

Pangerc et al. [8] noted that while most of the underwater noise thought to be associated with turbine operation is below 500 Hz, the band below 40 Hz is dominated by tidal flow-induced vibrations. Thus, in analogy to [8], this study is limited to analyzing and predicting PSDs in the frequency band between 40 and 520 Hz, considering only noise acquired when the turbine is rotating, i.e., for wind speeds between 3 and 24 m/s.

A. Introductory Analysis

The number of noise spectra in the data set as a function of wind speed, for the two turbines, is reported in Table I. The variability between spectra measured at a given wind speed is one of the analyses performed in [8]. As an example, Fig. 3(a)–(c) shows the noise PSDs recorded at turbine A5 for wind speeds: equal to 5 m/s; equal to 13 m/s; and ranging from 16 to 24 m/s. The related spectral averages are also shown.

The comparison of Fig. 3(a) and (b) confirms what was observed in [8]: the variability in PSDs measured at low speeds is considerable, and the tonal components are weak or dispersed over a band around 150 Hz; above 12 m/s, the variability decreases and the tonal components increase and are quite stable. In addition, it is shown in [8] that broadband acoustic intensity increases with increasing wind speed, from 3 to 10 m/s, while it remains almost stable at higher speeds. Based on these and other observations, Pangerc et al. [8] concluded that wind speed does not induce any important change in observed ambient

[8]The turbines are arranged on a grid whose nodes are roughly 1 km apart.

TABLE I
NUMBER OF PSDS COLLECTED IN [8] FOR TURBINES A1 AND A5, AS A FUNCTION OF THE WIND SPEED, FROM 3 TO 24 m/s

Wind speed [m/s]	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Number of PSDs for turbine A1	209	234	249	242	269	175	115	99	87	77	108	89	98	69	48	33	39	28	15	11	4	4
Number of PSDs for turbine A5	211	235	253	241	269	178	118	101	86	76	109	89	97	66	49	33	40	27	0	20	4	3

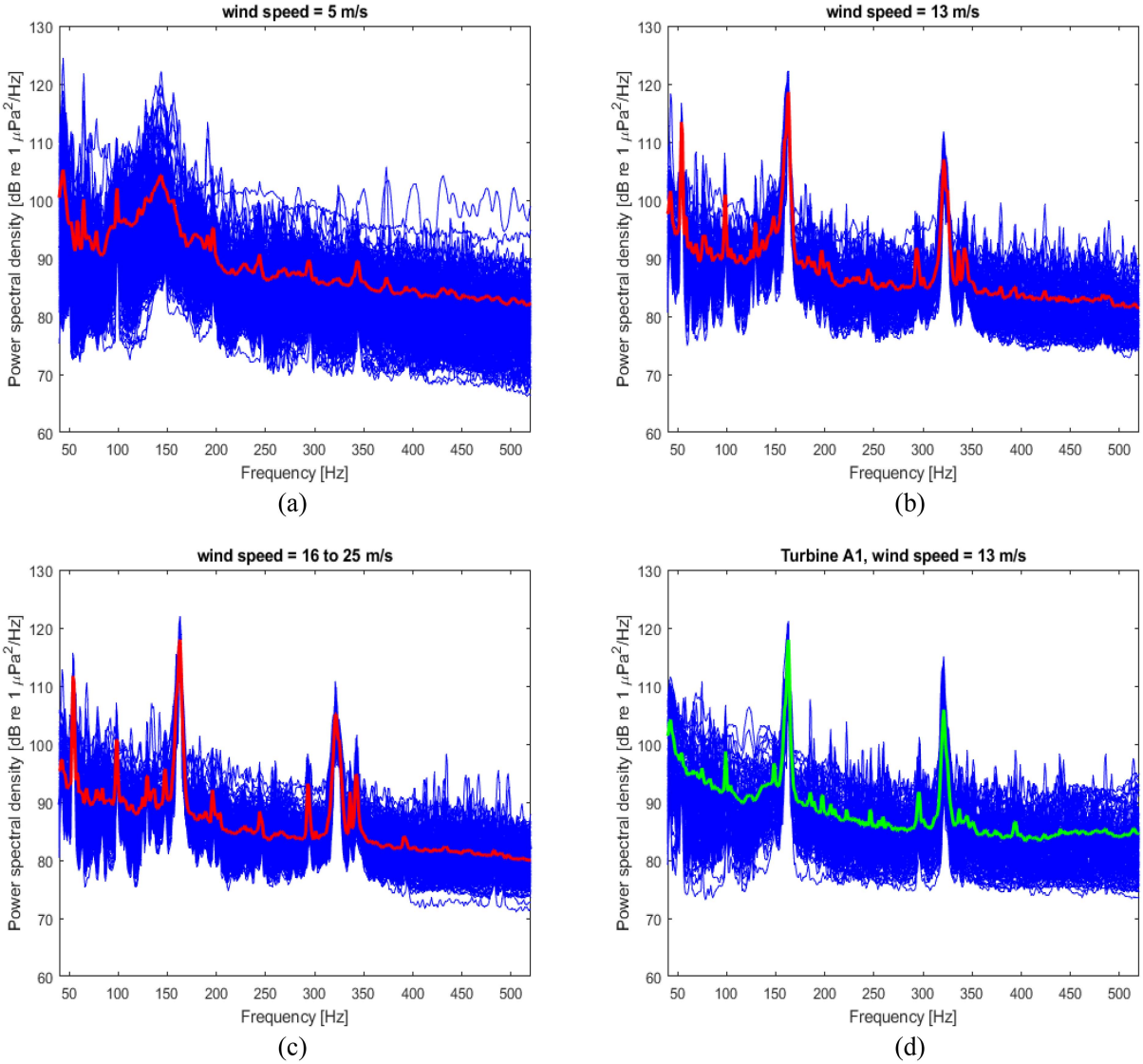


Fig. 3. PSDs collected at turbine A5 (blue lines) at a given wind speed with the related spectral averages (red lines): (a) 5 m/s; (b) 13 m/s; and (c) from 16 to 24 m/s, included. (d) PSDs collected at turbine A1 (blue lines) for a wind speed of 13 m/s with the related spectral average (green line).

noise,[9] and that the wind turbine does not contribute significantly to broadband noise, generating mostly tonal components.

Based on all the above, to counter the decreasing number of PSDs recorded for speeds greater than 15 m/s, in the rest of this article, noise spectra recorded at wind speeds from 16 to 24 m/s will be considered without distinction and will all be attributed (conventionally) to the speed of 16 m/s. Fig. 3(c) confirms the limited variability of the 242 PSDs recorded from 16 to 24 m/s, similar to that observed in Fig. 3(b) for the 109 PSDs recorded at 13 m/s.

A further observation concerns the similarity between the PSDs recorded at turbine A1 and those recorded at turbine A5. The comparison of Fig. 3(b) and (d) shows significant

[9]It is worth mentioning in this regard that acoustic measurements are made in very shallow water.

differences in the spectral lines, especially at 50, 100, and 350 Hz, confirming what was noticed in [8]: "differences are evident in the tonal characteristics of each of the turbines." Although the overall noise levels are similar, the above differences prevent the assumption that nominally identical turbines generate identical noise [8].

IV. Results and Discussion

A. Spectral Similarity Between Turbines A1 and A5

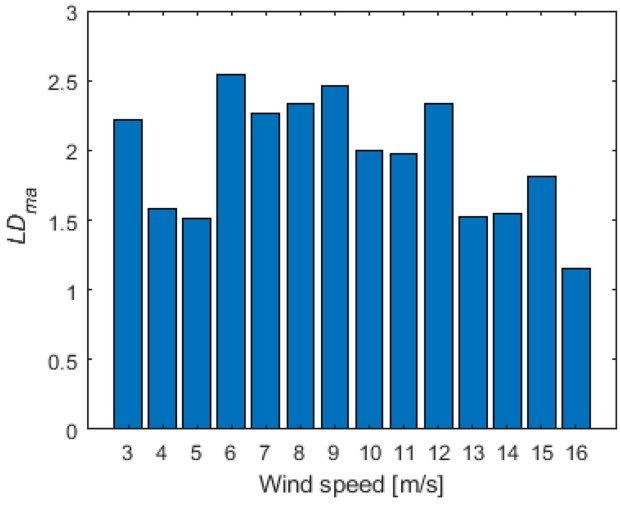
To set reference values for the three performance metrics defined in Section II-F, the first test is intended to measure the similarity between the spectral averages of all noise spectra collected at turbine A1, denoted $\mathbf{s}_{\mathrm{A1}}(w)$, and those of all spectra collected at turbine A5, denoted $\mathbf{s}_{\mathrm{A5}}(w)$, $w = 3,\ 4, \ldots,\ 16$ m/s.[10] As an example, the green line in Fig. 3(d) shows $\mathbf{s}_{\mathrm{A1}}(13)$, and the red line in Fig. 3(b) shows $\mathbf{s}_{\mathrm{A5}}(13)$. The histograms in Fig. 4 show the results obtained from (11)–(13), by using $\mathbf{s}_{\mathrm{A1}}(w)$ in place of $\hat{\mathbf{s}}(w)$ and $\mathbf{s}_{\mathrm{A5}}(w)$ in place of $\mathbf{s}(w)$. For the example in Fig. 3(b) and (d), the result is $\mathrm{LD}_{\mathrm{ma}}(13) = 1.53$ dB, $\mathrm{LD}_{\mathrm{rms}}(13) = 2.10$ dB, and $\mathrm{ISD}(13) = 0.19$. Averaging the three performance measures over the 14 wind speeds considered yields the following values: $\mathrm{LD}_{\mathrm{ma}} = 1.95$ dB, $\mathrm{LD}_{\mathrm{rms}} = 2.51$ dB, and $\mathrm{ISD} = 0.23$. By reversing prediction and spectral average (i.e., $\mathbf{s}_{\mathrm{A5}}(w)$ in place of $\hat{\mathbf{s}}(w)$ and $\mathbf{s}_{\mathrm{A1}}(w)$ in place of $\mathbf{s}(w)$), $\mathrm{LD}_{\mathrm{ma}}$ and $\mathrm{LD}_{\mathrm{rms}}$ remain unchanged, while the average of ISD becomes 0.17.[11]

B. Performance as a Function of the Number of Basis Spectra

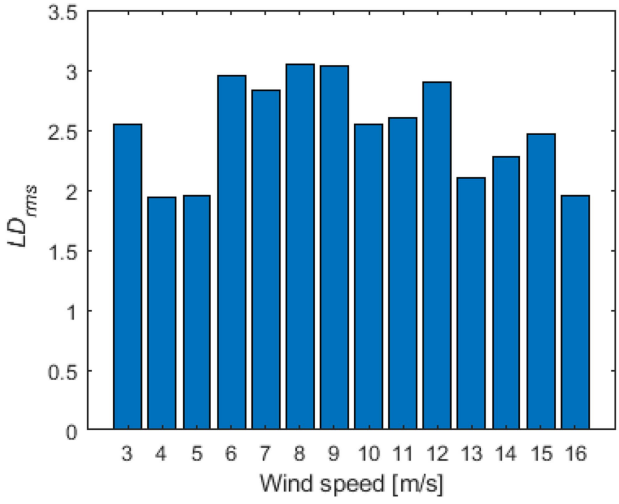
To compare the performance of prediction methods, it is necessary to set the numbers of basis spectra, q and r, used in the PCA and NMF analyses, respectively. For this purpose, the performance obtained as q and r varied was examined using the A5 turbine data in case study C1. For each value assigned to q and r, the prediction by PCA and NMF was repeated 100 times (randomly extracting the spectra that compose the *available set*), using CSS as the coefficient estimation technique. At each repetition, the three performance metrics were computed, for w from 3 to 16 m/s. At the end, the three metrics are averaged over wind speeds and repetitions, resulting in one value for each.

The values of the three metrics as q and r vary are shown in Fig. 5. While for PCA, $\mathrm{LD}_{\mathrm{ma}}$ and $\mathrm{LD}_{\mathrm{rms}}$ depend weakly on q, reaching the minimum for q between 3 and 5, for NMF, the two metrics show a significant minimum for $r = 20$. The ISD metric shows a less regular profile, but also in this case, small values are preferable for q (in particular, 2 and 4) and values around 25 are preferable for r. In light of these observations, in the rest of this article, the number of basis spectra in PCA will be $q = 4$ and that in NMF will be $r = 20$.

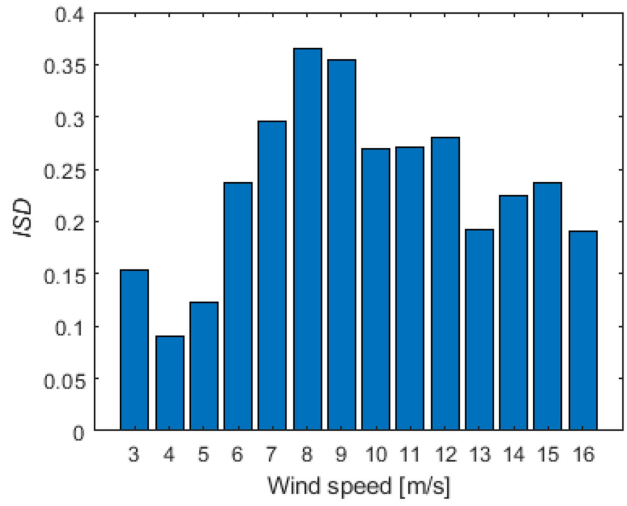
To have confirmation regarding this choice, the study was repeated in case study C3, using data from turbine A1 to compose the *available set* and those from turbine A5 to calculate the spectral averages [see Fig. 1(c)]. The coefficient estimation was performed by RFR. For PCA, the profiles of the three metrics



(a)



(b)

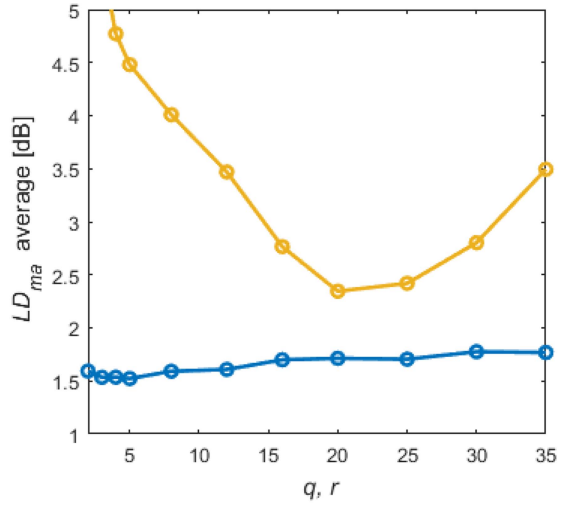


(c)

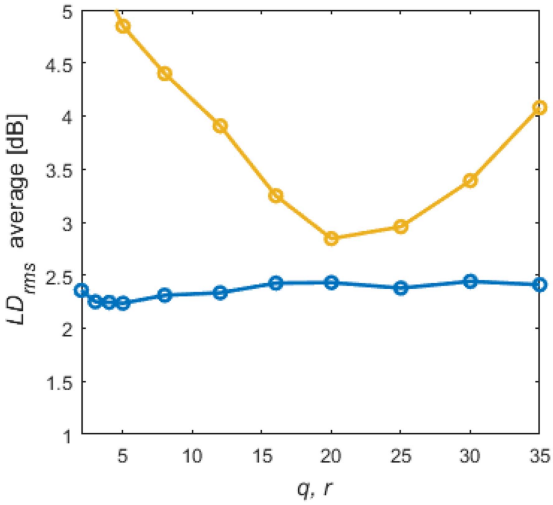
Fig. 4. Performance of predicting the spectral averages of noise measured at turbine A5 by using the spectral averages at turbine A1. (a) Mean absolute log-spectral distance; (b) root mean square log-spectral distance; (c) Itakura–Saito divergence.

[10] As discussed earlier, $w = 16$ m/s is used for all noise spectra recorded at speeds between 16 and 24 m/s, indistinctly.

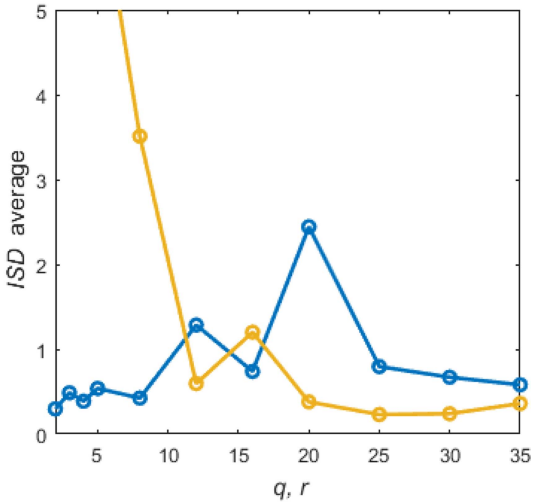
[11] ISD, not being a distance but a divergence, is asymmetrical.



(a)



(b)



(c)

Fig. 5. Average performance after 100 repetitions of case C1 with turbine A5 data, obtained by the PCA (blue dots) and NMF (golden dots), as a function of q and r. (a) Mean absolute log-spectral distance. (b) Root-mean-square log-spectral distance. (c) Itakura–Saito divergence.

TABLE II
PERFORMANCE OF PREDICTION METHODS AVERAGED OVER WIND SPEED; CASE STUDY C1, TURBINE A1, AND COEFFICIENT ESTIMATION BY GPR

	AAI	PCA	NMF
LD_{ma} [dB] average ± std. dev.	2.55 ± 1.00	1.38 ± 0.35	2.21 ± 0.63
LD_{rms} [dB] average ± std. dev.	3.13 ± 0.98	1.93 ± 0.41	2.85 ± 0.62
ISD average ± std. dev.	0.46 ± 0.41	0.18 ± 0.26	0.18 ± 0.08

For each metric, the green box shows the best prediction method, in terms of average and std. dev.

are very similar to those in Fig. 5, although the metric values are higher. To set $q = 4$ is again a reasonable choice. For NMF, the minima of $\mathrm{LD_{ma}}$ and $\mathrm{LD_{rms}}$ are for $r = 12$, while the minimum of ISD is for r between 16 and 25. Since for $r = 20$, the increases of $\mathrm{LD_{ma}}$ and $\mathrm{LD_{rms}}$ from their minima do not exceed 10%, overall $r = 20$ also proves to be a good choice for case study C3.

C. Comparison of Prediction Methods

The performance of the prediction methods described in Section II-B–II-D were compared with reference to case studies C1–C3. The combinations between the wind turbine (A1 or A5) and the coefficient estimation technique (see Section II-E) were changed from case to case, to avoid bias in the results and to extend the validity of the observations drawn. In all combinations, the prediction by AAI, PCA, and NMF was repeated 250 times (randomly extracting the spectra composing the *available set*), and in each repetition, wind speeds w from 3 to 16 m/s were considered. Some examples of the spectral predictions produced by the three methods, compared with spectral averages, are provided in Section IV-G.

For case study C1, wind turbine A1 and GPR estimation technique (setting the Matérn function with $\nu = 5/2$ as covariance kernel [14]) were selected. At each repetition, the *available set* consisted of 70 noise spectra, i.e., five spectra for each wind speed. Under these working conditions, AAI performed only averaging and no interpolation. Fig. 6 shows the average and std. dev. of the three metrics computed for each wind speed. Table II summarizes the values obtained when the three metrics are also averaged over wind speed.

For case study C2, wind turbine A5 and CSS estimation technique were selected. At each repetition, the *available set* consisted of 55 noise spectra, i.e., five spectra for each wind speed except 5, 9, and 10 m/s (spectra collected at these speeds are not present in the *available set*). Under these working conditions, AAI performed both averaging and interpolation. Fig. 7 shows the average and std. dev. of the three metrics computed for each wind speed. Table III summarizes the values obtained when the three metrics are also averaged over wind speed.

For case study C3, wind turbine A1 was used to compute the spectral predictions [i.e., turbine Φ in Fig. 1(c)], whereas turbine A5 was used to compute the spectral averages [i.e., turbine Γ

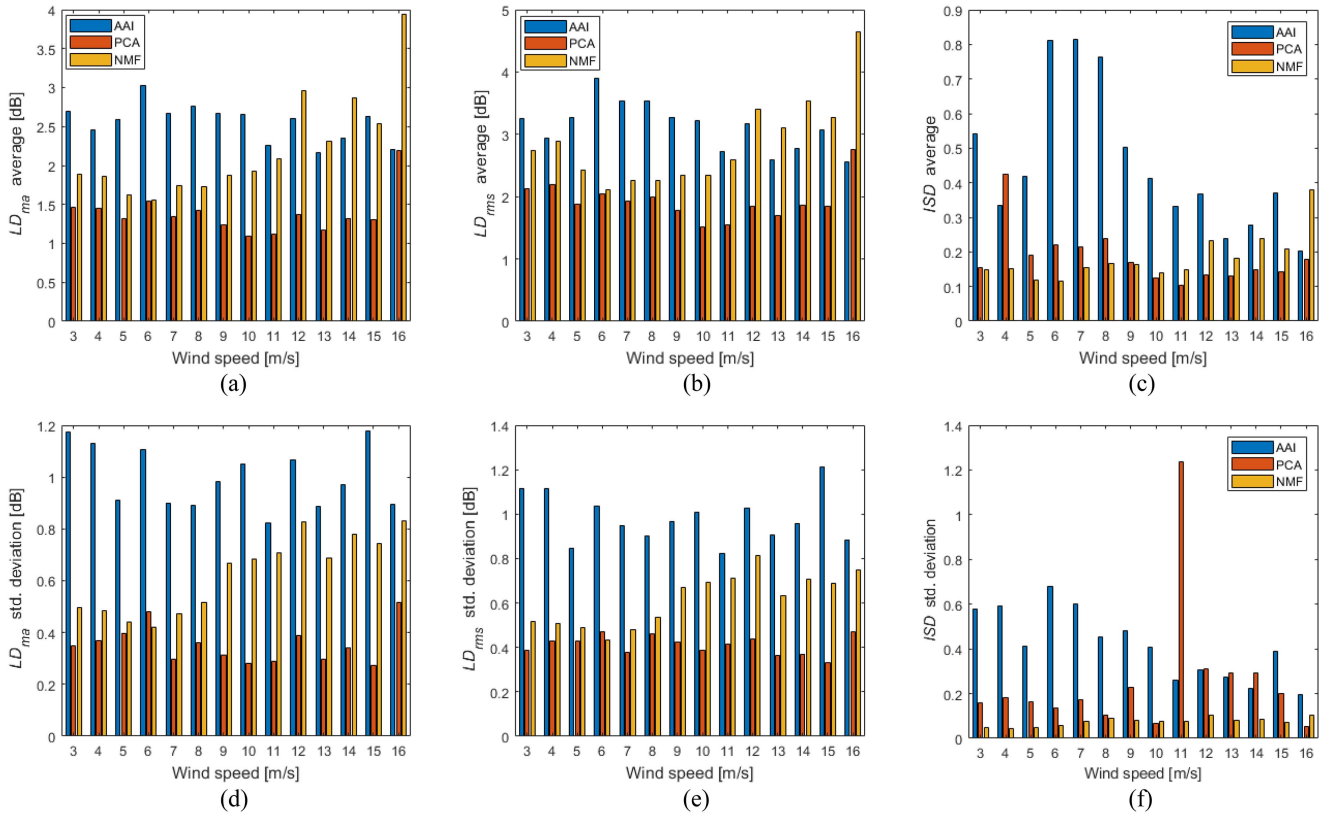


Fig. 6. Performance metrics as a function of wind speed and prediction method. Case study C1, turbine A1, and coefficient estimation by GPR were considered. The panels show the average and std. dev. of LD_{ma} in (a) and (d), LD_{rms} in (b) and (e), and ISD in (c) and (f). The legend is the same for all panels, even where absent.

TABLE III
PERFORMANCE OF PREDICTION METHODS AVERAGED OVER WIND SPEED; CASE STUDY C2 (*AVAILABLE SET* DEVOID OF SPECTRA AT THREE SPEEDS), TURBINE A5, AND COEFFICIENT ESTIMATION BY CSS

	AAI	PCA	NMF
LD_{ma} [dB] average ± std. dev.	2.18 ± 0.77	1.60 ± 0.42	2.30 ± 1.02
LD_{rms} [dB] average ± std. dev.	2.74 ± 0.80	2.32 ± 0.48	2.91 ± 1.05
ISD average ± std. dev.	0.34 ± 0.29	0.41 ± 1.43	0.47 ± 1.73

For each metric, the green box shows the best prediction method in terms of average and std. dev.

TABLE IV
PERFORMANCE OF PREDICTION METHODS AVERAGED OVER WIND SPEED; CASE STUDY C3 (*AVAILABLE SET* FROM TURBINE A1 AND SPECTRAL AVERAGES FROM TURBINE A5) AND COEFFICIENT ESTIMATION BY RFR

	AAI	PCA	NMF
LD_{ma} [dB] average ± std. dev.	2.90 ± 0.80	2.30 ± 0.38	3.05 ± 0.83
LD_{rms} [dB] average ± std. dev.	3.74 ± 0.76	3.12 ± 0.40	3.75 ± 0.80
ISD average ± std. dev.	0.73 ± 0.46	0.70 ± 1.42	0.41 ± 0.12

For each metric, the green box shows the best prediction method in terms of average and std. dev.

in Fig. 1(c)]. The RFR technique was selected for coefficient estimation, setting the main hyperparameters as follows: $\mathcal{E} = 30$ and $\ell = 8$. Fig. 8 shows the average and std. dev. of the three metrics computed for each wind speed. Table IV summarizes the values obtained when the three metrics are also averaged over wind speed.

From the analysis of the results reported in this section, some observations can be made.

1) For each prediction method, the oscillations of LD_{ma} and LD_{rms} as a function of wind are moderate, except for few special cases (e.g., NMF at $w = 16$ m/s in case C1, and NMF at $w = 3$ m/s in case C2). In contrast, the average of ISD shows noticeable oscillations, especially for AAI in case C1, PCA and NMF in case C2, and AAI and PCA in case C3.
2) At certain wind speeds, the ISD average and std. dev., for a given prediction method, are particularly high. This is

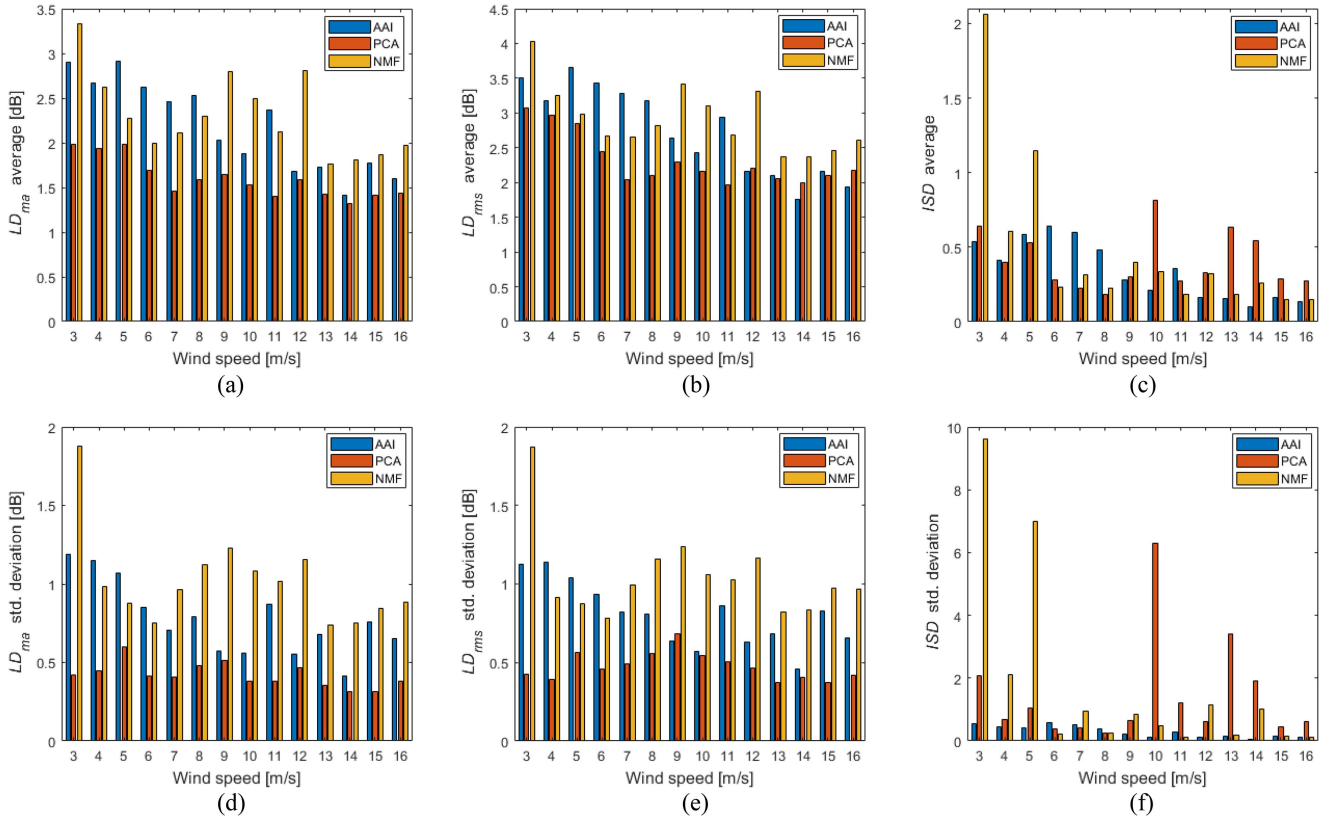


Fig. 7. Performance metrics as a function of wind speed and prediction method. Case study C2 (no spectra at 5, 9, and 10 m/s in the *available set*), turbine A5, and coefficient estimation by CSS were considered. The panels show the average and std. dev. of $\mathrm{LD_{ma}}$ in (a) and (d), $\mathrm{LD_{rms}}$ in (b) and (e), and ISD in (c) and (f).

especially observable in cases C2 and C3, for PCA and NMF, and means that within the 250 repetitions, there were several predictions for which ISD was very high.

3) In case C2, $\mathrm{LD_{ma}}$ and $\mathrm{LD_{rms}}$ metrics measured at the three speeds missing from the *available set* (i.e., 5, 9, and 10 m/s) are in line with those measured at the other speeds. This holds for both the average and the std. dev. In contrast, ISD values are particularly high for NMF at $w = 5$ m/s and for PCA at $w = 10$ m/s.
4) Moving from case C1 to case C2, Tables II and III show that for PCA and NMF, the three metrics worsen (in both average and std. dev.), while they improve for AAI. This discordant behavior may be related to the fact that both the turbine and coefficient estimation technique changed in moving from C1 to C2.
5) Moving from case C1 to case C3, Tables II and IV show that the averages of the three metrics worsen for all prediction methods, while the std. dev. values are quite stable (except ISD std. dev. for PCA).
6) The green boxes in Tables II–IV indicate, for each metric, the prediction method that performs best in terms of both average and std. dev. For $\mathrm{LD_{ma}}$ and $\mathrm{LD_{rms}}$, the best prediction method is always PCA, with significantly better values than those reported by AAI and NMF. Figs. 6–8 show that PCA average and std. dev. are the lowest at all wind speeds, except for very few cases, where they are still close to the best. The ISD metric, on the other hand, is never favorable to PCA; it rewards NMF prediction in cases C1 and C3, and AAI prediction in case C2. Finally, it is observed that the NMF prediction performs better than AAI in case C1, worse in case C2, and almost equivalent in case C3.

D. Varying Number of Available Spectral Samples

In this study, the number of available spectra is generally set equal to 5 for each wind speed. This section evaluates the impact on the results of a random number of available spectra by fixing a discrete random variable for each speed, ranging from 1 to 9, with uniform density. Although the mean number is always 5, in this way, the actual number of available spectra fluctuates considerably depending on the speed and repetition of the experiment.

Table V shows the results obtained in case study C1 with the same turbine and estimation technique adopted in the previous section (i.e., A1 and GPR), but with an *available set* consisting of a number of spectral samples that varies from one repetition of the experiment to the next (70 being the mean number). Comparing Table V with Table II shows an increase in all metric values, both averages and std. dev. However, for the averages of $\mathrm{LD_{ma}}$ and $\mathrm{LD_{rms}}$, the increase is about 10% for AAI, 2% for PCA, and 5% for NMF. Even for the std. dev., the smallest increase is for PCA, which is less than 20%. Finally, the increase in the average of ISD is about 78% for AAI, 17% for PCA, and 6% for NMF.

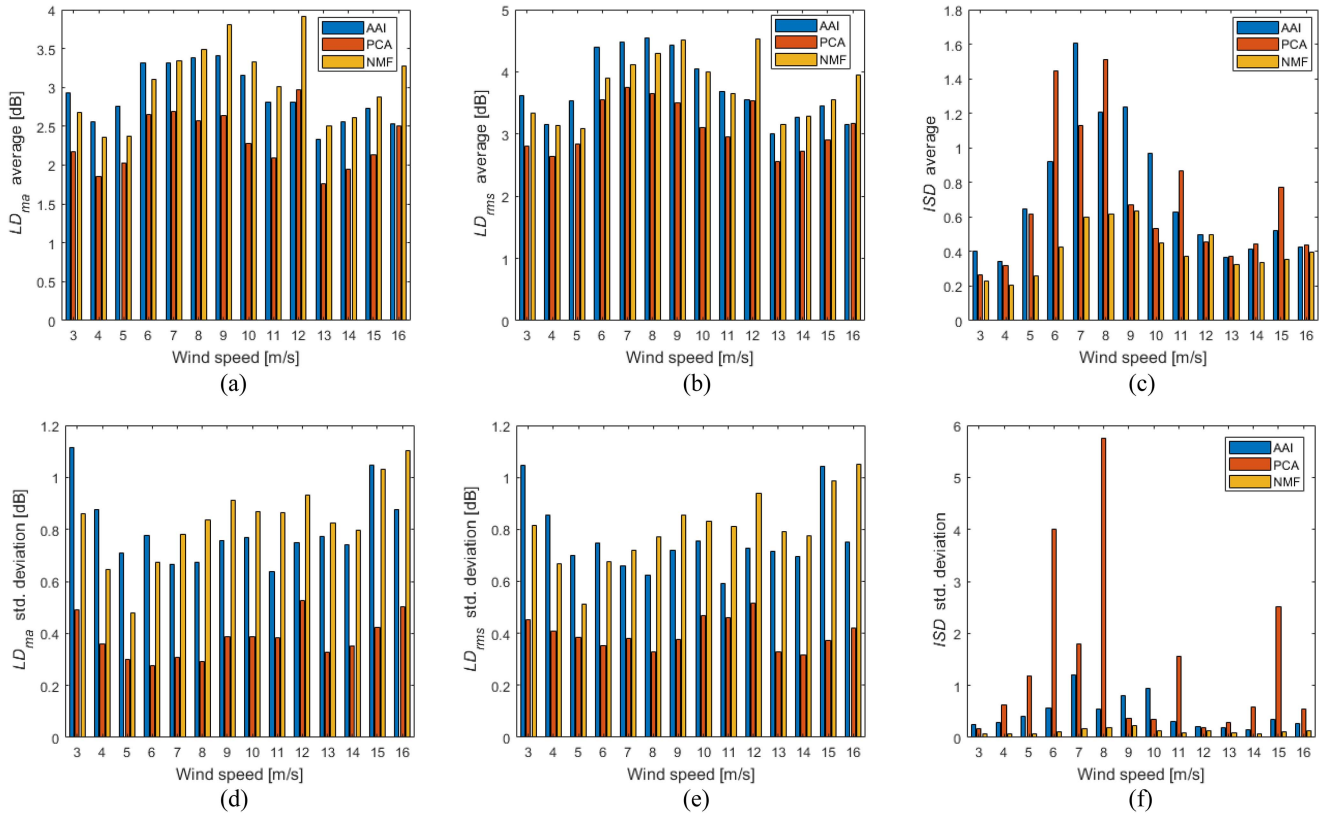


Fig. 8. Performance metrics as a function of wind speed and prediction method. Case study C3 (*available set* from turbine A1 and spectral averages from turbine A5) and coefficient estimation by RFR were considered. The panels show the average and std. dev. of LD_{ma} in (a) and (d), LD_{rms} in (b) and (e), and ISD in (c) and (f). The legend is the same for all panels, even where absent.

TABLE V
PERFORMANCE OF PREDICTION METHODS AVERAGED OVER WIND SPEED; CASE STUDY C1 WITH RANDOM NUMBER OF SPECTRA COMPOSING THE *AVAILABLE SET* AT EACH SPEED (RANGING FROM 1 TO 9), TURBINE A1 DATA, AND COEFFICIENT ESTIMATION BY GPR

	AAI	PCA	NMF
LD_{ma} [dB] average ± std. dev.	2.83 ± 1.51	1.41 ± 0.41	2.32 ± 0.81
LD_{rms} [dB] average ± std. dev.	3.43 ± 1.55	1.97 ± 0.49	2.97 ± 0.79
ISD average ± std. dev.	0.82 ± 1.61	0.21 ± 0.72	0.19 ± 0.10

For each metric, the green box shows the best prediction method in terms of average and std. dev.

The imbalance in the *available set*, due to the variable number of spectra among wind speeds, impacts the variances of the metrics more than their averages. On the other hand, the worsening observed for the PCA-based method are very limited, especially those affecting the average values of LD_{ma} and LD_{rms}.

E. Comparison of Coefficient Estimation Techniques

In Section IV-C, the PCA-based prediction method is the one that, with a few exceptions, performed better and with less variance. This evaluation is further confirmed in the test described in Section IV-D. This section compares the performance that the fitting and regression techniques described in Section II-E achieve in case studies C1–C3, when used in combination with the PCA-based prediction. For all the case studies, the assessment was repeated 250 times (randomly extracting the spectra that compose the *available set*), and at each repetition, the performance achieved by PD2, CSS, RFR, and GPR was registered. The same parameters mentioned earlier were adopted: $q = 4$, $\mathcal{E} = 30$, $\ell = 8$, and Matérn function with $\nu = 5/2$. Different settings of the RFR hyperparameters (i.e., $\mathcal{E}$ and ℓ) and the GPR kernel are discussed in Sections C and D of Appendix A, respectively.

For case study C1, data from wind turbine A1 and *available sets* consisting of 70 noise spectra (i.e., five samples for each wind speed) were used. For case C2, data from turbine A5 and *available sets* consisting of five spectra for each wind speed except 4, 7, 12, and 15 m/s were used (50 samples in total). For case C3, turbine A5 was used for computing the spectral predictions, whereas turbine A1 was used to compute the spectral averages.[12]

[12]To enrich the combinations tested, this choice is the opposite of what was done for case C3 in Section IV-C.

TABLE VI
PERFORMANCE OF ESTIMATION TECHNIQUES AVERAGED OVER WIND SPEED; CASE STUDY C1, TURBINE A1, AND PCA-BASED PREDICTION

	PD2	CSS	RFR	GPR
LD_{ma} [dB] average ± std. dev.	1.42 ± 0.35	1.41 ± 0.41	1.41 ± 0.42	1.37 ± 0.37
LD_{rms} [dB] average ± std. dev.	2.04 ± 0.40	2.02 ± 0.49	2.01 ± 0.50	1.92 ± 0.43
ISD average ± std. dev.	0.22 ± 0.63	0.43 ± 2.00	0.40 ± 1.91	0.18 ± 0.50

For each metric, the light blue box shows the best estimation method, in terms of average.

TABLE VII
PERFORMANCE OF ESTIMATION TECHNIQUES AVERAGED OVER WIND SPEED; CASE STUDY C2 (*AVAILABLE SET* DEVOID OF SPECTRA AT FOUR SPEEDS), TURBINE A5, AND PCA-BASED PREDICTION

	PD2	CSS	RFR	GPR
LD_{ma} [dB] average ± std. dev.	1.90 ± 0.44	1.69 ± 0.49	1.70 ± 0.49	1.72 ± 0.43
LD_{rms} [dB] average ± std. dev.	2.80 ± 0.44	2.41 ± 0.52	2.40 ± 0.53	2.47 ± 0.47
ISD average ± std. dev.	0.28 ± 0.24	0.42 ± 1.17	0.29 ± 0.42	0.18 ± 0.16

For each metric, the light blue box shows the best estimation method, in terms of average.

TABLE VIII
PERFORMANCE OF ESTIMATION TECHNIQUES AVERAGED OVER WIND SPEED; CASE STUDY C3 (*AVAILABLE SET* FROM TURBINE A5 AND SPECTRAL AVERAGES FROM TURBINE A1) AND PCA-BASED PREDICTION

	PD2	CSS	RFR	GPR
LD_{ma} [dB] average ± std. dev.	2.48 ± 0.33	2.35 ± 0.38	2.36 ± 0.39	2.33 ± 0.34
LD_{rms} [dB] average ± std. dev.	3.33 ± 0.34	3.10 ± 0.37	3.09 ± 0.38	3.06 ± 0.33
ISD average ± std. dev.	0.46 ± 1.31	0.59 ± 2.60	0.49 ± 1.30	0.38 ± 1.06

For each metric, the light blue box shows the best estimation method, in terms of average.

Tables VI–VIII report the average and std. dev. of the three metrics for cases C1–C3, respectively, following additional averaging over wind speed. The values of the three metrics for each wind speed are particularly significant in case C2 (where four speeds are not represented in the *available set*) and are shown in Fig. 9.

From the analysis of the results reported in this section, some observations can be made.

1) In case study C1, the four coefficient estimation techniques provide nearly equivalent LD_{ma} and LD_{rms} metrics, both in terms of average and std. dev. In cases C2 and C3, this holds for CSS, RFR, and GPR, while the performance of PD2 is worse.
2) In the three case studies considered, the ISD metric always identifies the best and worst technique, in terms of both average and std. dev. The best is GPR; the worst is CSS.
3) In case study C2, LD_{ma} and LD_{rms} metrics shown in Fig. 9 confirm the similarity between the performance of CSS, RFR, and GPR. At some wind speeds, the average PD2 is significantly worse. At some speeds, the ISD metrics are considerably worse than at other speeds. This is especially true for PD2 and CSS, to a lesser extent for RFR. Performance at speeds that are not present in the *available set* is in line with those next to them. The only exception is 12 m/s, at which all coefficient estimation techniques show an increase in both LD_{ma} and LD_{rms}.
4) The PCA columns in Tables II–IV can be compared with GPR, CSS, and RFR columns in Tables VI–VIII, respectively. Although in case C2, the speeds that are not present in the *available set* change, and for case C3, the turbines used for predicting and spectral averaging are reversed; the performance obtained is very close. An exception to this is the ISD metric, which shows some fluctuations, especially in the std. dev. values.

F. Cross-Prediction Performance

Section IV-A reported the performance obtained when, at each wind speed, the average of all noise spectra collected at a given turbine is used to predict the spectral average at the other turbine. Table IX summarizes these results and compares them with those obtained using only five noise spectra collected at a given turbine to predict, through PCA and GPR, the spectral average at the other turbine (i.e., case study C3). In the latter case, 250 repetitions are used to compute average and std. dev. of the metrics.

It can be observed that moving from prediction that exploits all spectra to prediction that uses only five spectra for each wind speed (and leverages PCA and GPR), performance deteriorates, but this occurs to a limited extent. LD_{ma} worsens on average by 0.40 dB, whereas LD_{rms} worsens on average by 0.61 dB. ISD values do not exceed 0.44 on average, but their variance is considerable.

G. Examples of Noise Spectral Prediction

This section shows some examples of prediction obtained by the methods considered. The purpose is to give a visual impression of the difference between possible outcomes and their respective performance metrics. Since these examples constitute individual realizations of random processes, it is not possible to ascribe general or statistical meaning to them.

The spectral average of turbine A5 at wind speed of 13 m/s, already shown by the red line in Fig. 3(b), is chosen as the prediction target. In Section IV-A, an attempt was made to predict such a spectral average by averaging all noise spectra recorded at turbine A1 at the same wind speed. The result was shown by the green line in Fig. 3(d), and the performance metrics are copied in Table X, for comparison. Figs. 10–12 show the prediction target (brown lines) and a realization of the predictions (blue lines) obtained by the methods considered: AAI, PCA, and NMF. In the last two cases, coefficient estimation was performed by the GPR technique. Fig. 10 is about case study C1. Fig. 11 is about case study C2, in which the *available set* is lacking spectra collected at

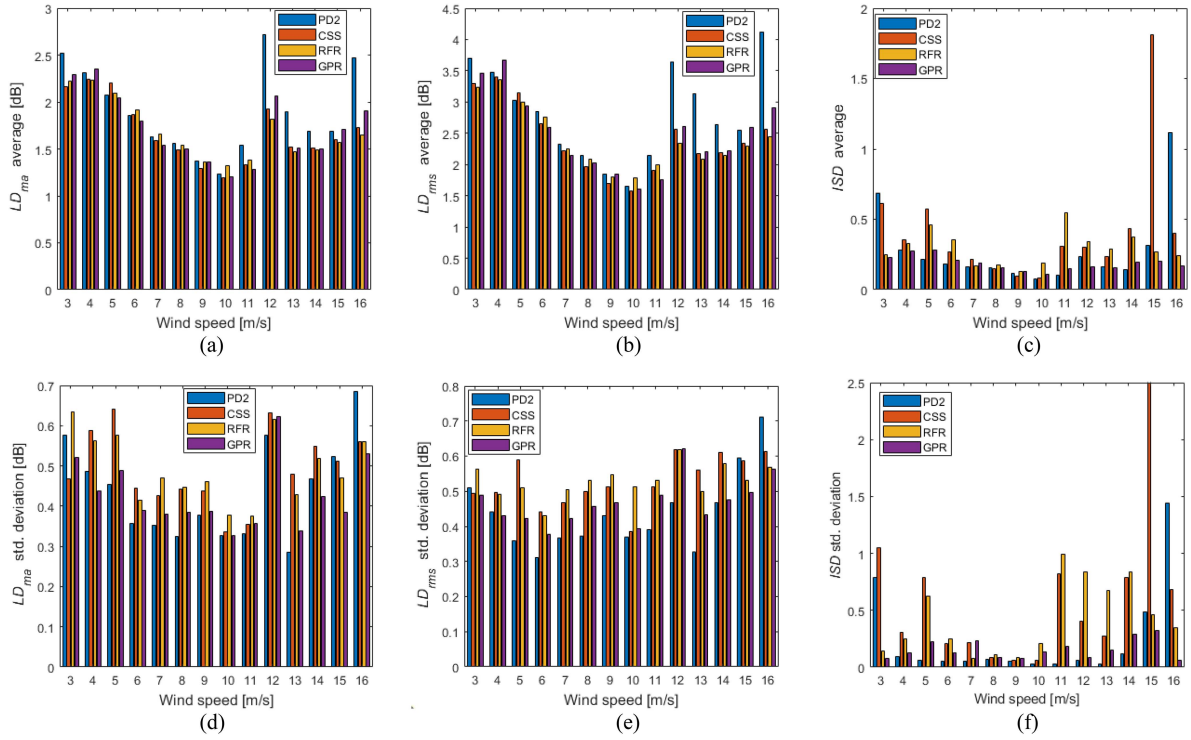


Fig. 9. Performance metrics as a function of wind speed and coefficient estimation technique. Case study C2 (no spectra at 4, 7, 12, and 15 m/s in the *available set*), turbine A5, and PCA-based prediction are considered. The panels show the average and std. dev. of LD_{ma} in (a) and (d), LD_{rms} in (b) and (e), and ISD in (c) and (f). In (f), ISD std. dev. at 15 m/s for the CSS technique is equal to 10.67.

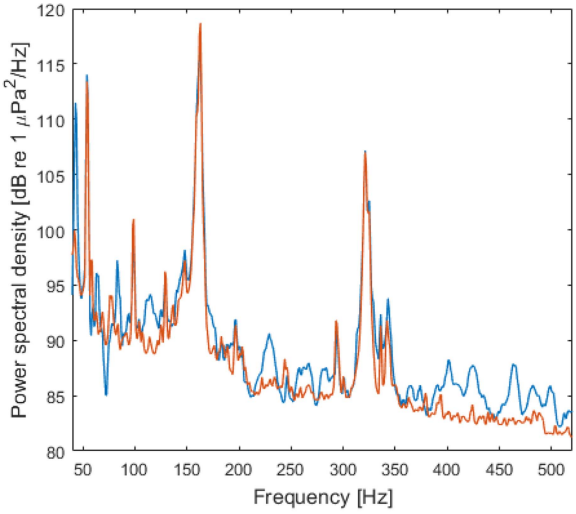
TABLE IX
PERFORMANCE IN PREDICTING THE SPECTRAL AVERAGES OF A GIVEN TURBINE BY AVERAGING ALL THE SPECTRA OF THE OTHER TURBINE OR USING ONLY FIVE SPECTRA OF IT (PROCESSED BY PCA AND GPR)

	Spectral averages of turbine A1 are predicted using data from turbine A5		Spectral averages of turbine A5 are predicted using data from turbine A1	
For each wind speed:	*all A5 spectra are averaged*	*five A5 spectra are used by PCA and GPR*	*all A1 spectra are averaged*	*five A1 spectra are used by PCA and GPR*
LD_{ma} [dB] average ± std. dev.	1.95	2.33 ± 0.34	1.95	2.35 ± 0.37
LD_{rms} [dB] average ± std. dev.	2.51	3.06 ± 0.33	2.51	3.12 ± 0.37
ISD average ± std. dev.	0.17	0.38 ± 1.06	0.23	0.44 ± 0.46

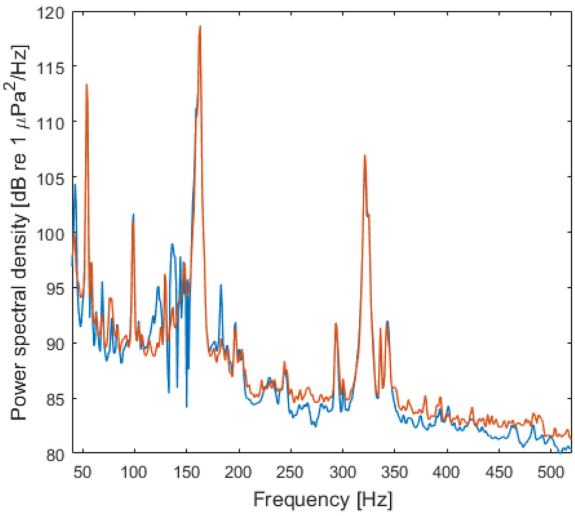
TABLE X
PREDICTION OF THE SPECTRAL AVERAGE AT $w = 13$ m/s FOR TURBINE A5; PERFORMANCE OF A SINGLE EXAMPLE IN CASE STUDIES C1, C2 (SPEEDS NOT INCLUDED IN THE *AVAILABLE SET*: 4, 8, 10, AND 13 m/s), AND C3

	C1			C2			C3			Average of all A1 spectra at w = 13 m/s
	AAI	PCA	NMF	AAI	PCA	NMF	AAI	PCA	NMF	
Fig.	10(a)	10(b)	10(c)	11(a)	11(b)	11(c)	12(a)	12(b)	12(c)	3(b), 3(d)
LD_{ma} [dB]	1.85	1.15	1.55	2.26	1.20	2.60	2.69	1.70	3.23	1.53
LD_{rms}[dB]	2.39	1.63	2.94	2.94	1.98	3.51	3.71	2.33	3.84	2.10
ISD	0.12	0.08	0.13	0.17	0.27	0.21	0.30	0.23	0.35	0.19

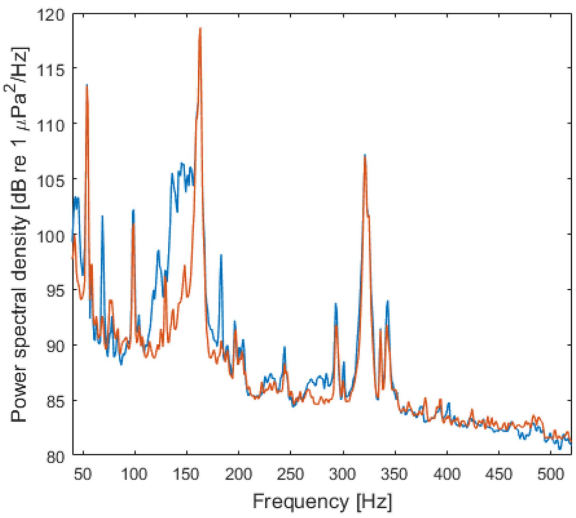
GPR estimation was used with PCA and NMF. The last column shows the performance obtained using the average of all noise spectra at w = 13 m/s collected at turbine A1, as reported in Section IV-A.



(a)

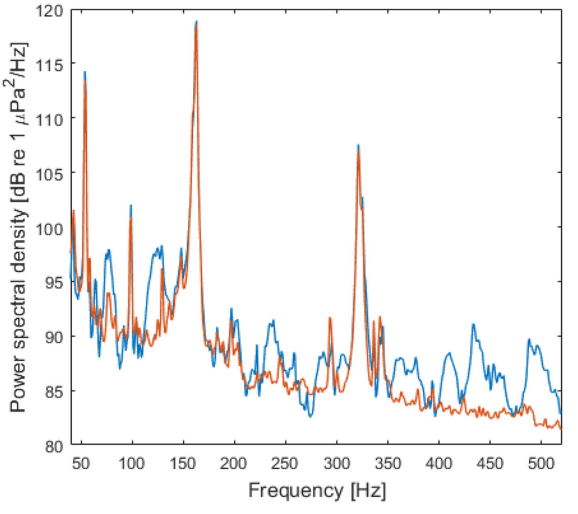


(b)

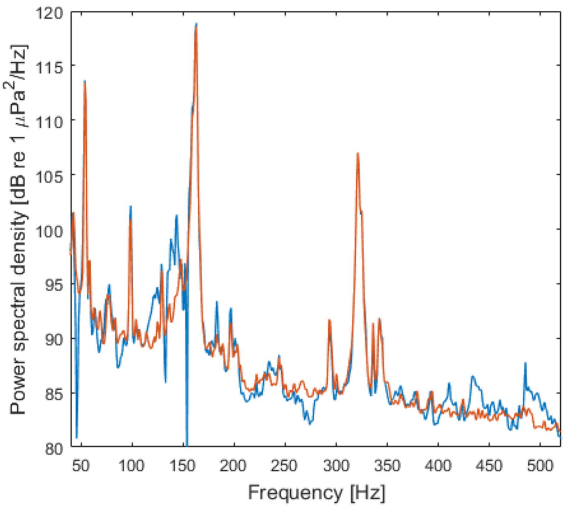


(c)

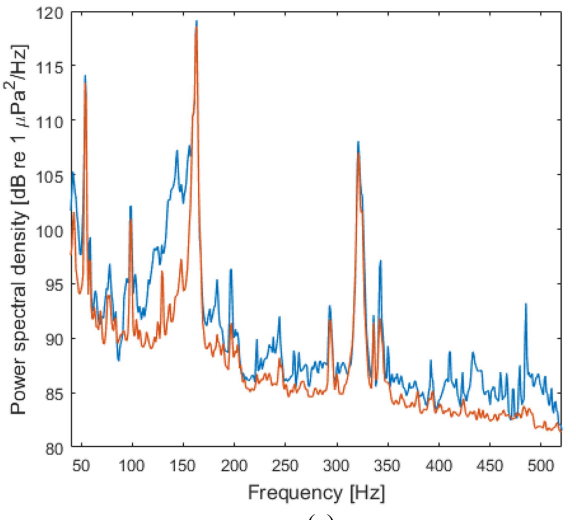
Fig. 10. Prediction of the spectral average at 13 m/s for turbine A5 (brown line) in case study C1. Example of result (blue line) obtained by (a) AAI, (b) PCA and GPR, and (c) NMF and GPR.



(a)



(b)



(c)

Fig. 11. Prediction of the spectral average at 13 m/s for turbine A5 (brown line) in case study C2, not including the speeds 4, 8, 10, and 13 m/s in the *available set*. Example of result (blue line) obtained by (a) AAI, (b) PCA and GPR, and (c) NMF and GPR.

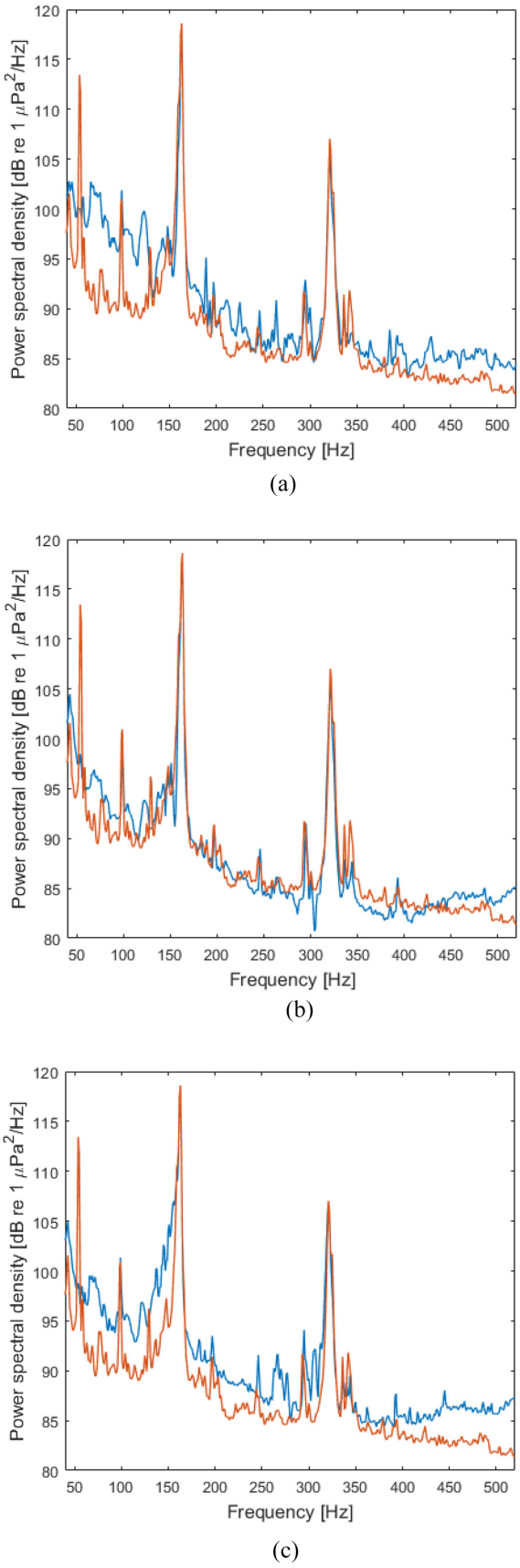


Fig. 12. Prediction of the spectral average at 13 m/s for turbine A5 (brown line) in case study C3. Example of result (blue line) obtained by (a) AAI, (b) PCA and GPR, and (c) NMF and GPR.

four speeds, including that of the prediction target (i.e., 13 m/s). Finally, Fig. 12 is about case C3. For each prediction example, performance metrics are calculated and collected in Table X.

For case C1, Table X shows that PCA prediction has better metrics than AAI and NMF predictions. Comparison of the spectra in Fig. 10 shows that, on the one hand, PCA prediction reproduces broadband noise more accurately than AAI and NMF; on the other hand, it reproduces the target spectral lines more precisely, introducing fewer spurious lines than the other prediction methods. However, in a small band around 150 Hz, PCA prediction shows significant oscillations, with narrow and deep local minima. Moving from case C1 to case C2, all metrics worsen, but the advantages of PCA prediction described for case C1 are found, amplified, for case C2 as well (see Fig. 11). However, the ISD value for PCA prediction is now the worst of the three. This is probably due to the two local very deep minima at about 50 and 150 Hz [see Fig. 11(b)]. Finally, the use of five noise spectra collected at turbine A1, for case C3, leads to a deterioration in the reproduction of the target spectral lines (see Fig. 12, especially the lines at about 50 and 350 Hz), which in turn worsens the metrics. Nevertheless, the advantage of PCA prediction is clearly visible and confirmed by the metric values. Table X shows that the result obtained by PCA for case C3, using only five spectra collected at turbine A1, is only slightly worse than that obtained by averaging all spectra recorded at 13 m/s (i.e., 108 spectra, according to Table I).

V. Conclusion

Underwater acoustic noise recorded near a wind turbine in very shallow water depends on wind speed: as the speed varies, the spectral lines generated by the turbine vary in frequency and amplitude. The first question that this study sought to answer is how to approximate the spectral average of the underwater noise, at a given wind speed, having only a few noise samples (each sample consisting of the PSD obtained over a 10-min recording interval) collected at that speed. A problem variation is to have no samples collected at that speed, having only a few samples collected at higher and lower speeds.

This study demonstrated that, in both the cases, it is possible to obtain better performance than that provided by AAI operations, particularly by performing PCA of the available samples, followed by a coefficient estimation step performed by a fitting or regression technique. In contrast, the analysis by NMF did not produce performance improvements that were general enough. The fitting and regression techniques tested did not show significant differences in performance, although regression based on supervised learning was slightly advantageous compared to curve fitting. In particular, GPR was found to be the most suitable estimation technique to be combined with PCA.

It should be noted that these conclusions are based on the analysis of the two data sets provided in [8], each related to a given turbine and composed of 76–270 underwater noise PSDs for each wind speed between 3 and 16 m/s, in steps of 1 m/s. Focusing on a specific turbine, by adopting PCA with GPR and having five PSDs available for each wind speed, the noise spectral average is predicted with an LD_{ma} of about 1.4 $\pm$

0.4 dB, whereas if there are no PSDs available at four wind speeds, $\mathrm{LD_{ma}}$ rises to about 1.7 ± 0.5 dB. The aim of this study was to find a prediction pipeline robust enough to produce good results in all the case studies tested, not to optimally set the parameters involved. It cannot be excluded that a more accurate algorithm selection and parameter setting could further improve performance.

In [8], it was observed that the spectral averages of the noise recorded at two nominally identical turbines of the same OWF show considerable differences. This is confirmed by $\mathrm{LD_{ma}}$ measured in this study, equal to 1.95 dB. Consequently, the second question this study sought to answer is how to approximate the spectral average of the noise recorded at a given turbine, having only a few noise samples recorded at another (nominally identical) turbine, at the same wind speed. Again, prediction based on PCA performed better than prediction based on mean and interpolation or NMF. The use of PCA with GPR and only five spectra for each wind speed allows one to approach the results obtained by averaging hundreds of spectra: $\mathrm{LD_{ma}}$ is about 2.35 ± 0.35 dB.

The considerable difference between the noises observed at the two turbines leads to a further conclusion: it is advantageous to predict the noise spectral average at a given turbine by using a few spectra (namely, five PSDs for each wind speed) recorded at the turbine itself ($\mathrm{LD_{ma}} = 1.4 \pm 0.4$ dB) rather than hundreds of spectra recorded at another turbine ($\mathrm{LD_{ma}} = 1.95$ dB). This conclusion is strictly related to the data sets considered: future investigations using other data sets are necessary to establish its generality.

Appendix A

In the problem at hand, a regression model should be trained for each coefficient to be estimated, generally indicated by y. The model is a function $f, \hat{y} = f(w)$, that assigns the coefficient value $\hat{y}$ at the wind speed w. The independent variable is just one: w. The training set is composed of the n values of the coefficient y, for each wind speed w_j, $j \in [1, m]$, obtained from the analysis of the *available set* (an example is shown in Fig. 2). The training set is composed of $N = mn$ samples, each of them is a couple (w_j, y). A test sample used in the prediction phase is simply an arbitrary value of w at which the coefficient value $\hat{y}$ should be estimated. Therefore, the number of features is limited to one: the wind speed w.

A. Random Forest Regression

Random forest [11], [20] is a tree-based ensemble method, which combines bagging and random attribute subset selection, adopting classification and regression trees (CARTs) as weak learners. According to the bagging philosophy, each weak learner is trained using a randomly sampled subset of the whole training set. In general, partitioning at each node is performed by considering only a random subset of the features: however, this has no impact in this specific case, since the number of features is limited to one.

Concerning the weak learner, each CART is an acyclic connected graph, where each node represents a decision rule (called split) that is related to a single feature and that leads to partitioning the data in two groups. During the training phase, each tree is automatically grown by iteratively identifying the feature that yields the best split in terms of a preselected metric (e.g., the Gini index for classification and the mean squared error for regression) [11], [20]. At the root node, the samples of the training subset are split for the first time. Then, the resulting two groups are forwarded down the tree. Each subsequent node performs further splitting of those training samples that come from the father node. Such a process is ended when the group of samples assigned to a node satisfies a termination condition (in the problem at hand, a split node is considered if it leaves at least ℓ training samples in each of the child nodes). During the prediction based on a new sample (i.e., the test phase), the sample traverses the tree by following the rules learned during the training phase. Once it reaches the terminal node (called leaf), the output is computed based on the statistics of the set of training samples assigned to that leaf. In case of regression, the output $\hat{y}$ corresponds to the average of the y values of such samples.

Let $\mathcal{E}$ denote the number of trees in the ensemble, and let $\mathcal{T}_\varepsilon$, $\varepsilon \in [1, \mathcal{E}]$, identify the εth tree. The RFR training procedure for the problem at hand can be summarized according to the scheme in Table XI.

Concerning the prediction phase, each new sample w is processed by each one of the $\mathcal{E}$ trees composing the forest. For each one of such trees, a prediction $\hat{y}_\varepsilon$ is obtained, as detailed earlier. The final prediction is then obtained by averaging the $\mathcal{E}$ predictions in the forest.

The results reported in this article were obtained with $\mathcal{E} = 30$ and $\ell = 8$. Section C of this appendix discusses the issues involved in optimizing these hyperparameters.

B. Gaussian Process Regression

GPR [14], [21] assumes that a stochastic function f maps the input $\boldsymbol{\psi}$ to the output y according to the relation $y = f(\boldsymbol{\psi}) + \delta$, where δ is a noise term, with a normal distribution $\mathcal{N}(0, \sigma_\delta^2)$, which models the intrinsic randomness in the observations. In GPR, the function $f(\boldsymbol{\psi})$ is assumed to be distributed as a Gaussian process $\mathcal{GP}(m(\boldsymbol{\psi}), k(\boldsymbol{\psi}, \boldsymbol{\psi}'))$. While a multivariate Gaussian distribution is specified by its mean and covariance, which are a vector and a matrix, respectively, a Gaussian process is specified by its mean function, $m(\boldsymbol{\psi})$, and covariance function, $k(\boldsymbol{\psi}, \boldsymbol{\psi}')$. Therefore, the function f is a Gaussian process, which, in turn, is a distribution over functions [14]. The covariance function measures the joint variability of the function values at different inputs, $\boldsymbol{\psi}$ and $\boldsymbol{\psi}'$, and is commonly called the *kernel* of the Gaussian process. The choice of kernel allows reasonable assumptions to be made about the smoothness and recurrent patterns expected in the data.

In general, let $\boldsymbol{\Psi}$ be a matrix in which each row is an input $\boldsymbol{\psi}$, let $\mathbf{y}$ be the column vector of the corresponding outputs, and let $\mathbf{f}_*$ be a vector of samples values of the function f at inputs $\boldsymbol{\Psi}_*$. When a set $\mathcal{S}_t = \{\boldsymbol{\Psi}_t, \mathbf{y}_t\}$ of observations is available, the prior knowledge about the stochastic function f can be updated in light of the training data. As shown in [21], setting

TABLE XI
RFR TRAINING PROCEDURE FOR THE PROBLEM AT HAND

Hyperparameters: the number of trees, $\mathcal{E}$, and the minimum number of samples to be at a leaf node, ℓ.
Sampling: randomly sample the training set to create $\mathcal{E}$ subsets of the training data, each containing the same number of samples as the training set. Such sampling is random with replacement, hence implying that some samples may be taken more than once and other samples may be never used.
Training: repeat for each tree $\mathcal{T}_\varepsilon$, $\varepsilon = 1, 2, \ldots, \mathcal{E}$: • For a given tree node, identify the best decision rule to split the node into two child nodes. • Repeat the point above until the number of samples assigned to the child nodes is greater than or equal to ℓ.

TABLE XII
RFR PERFORMANCE VARIATION WITH $\mathcal{E}$ AND ℓ; CASE STUDY C1, TURBINE A1, AND PCA-BASED PREDICTION, AS IN TABLE VI

PCA and RFR	$\mathcal{E} = 30$ $\ell = 8$	$\mathcal{E} = 30$ $\ell = 16$
LD_{ma} [dB] average ± std. dev.	1.41 ± 0.42	1.38 ± 0.39
LD_{rms} [dB] average ± std. dev.	2.01 ± 0.50	1.91 ± 0.45
ISD average ± std. dev.	0.40 ± 1.91	0.22 ± 0.69

TABLE XIII
RFR PERFORMANCE VARIATION WITH $\mathcal{E}$ AND ℓ; CASE STUDY C2, TURBINE A5, AND PCA-BASED PREDICTION, AS IN TABLE VII

PCA and RFR	$\mathcal{E} = 30$ $\ell = 8$	$\mathcal{E} = 30$ $\ell = 16$
LD_{ma} [dB] average ± std. dev.	1.70 ± 0.49	1.74 ± 0.41
LD_{rms} [dB] average ± std. dev.	2.40 ± 0.53	2.54 ± 0.44
ISD average ± std. dev.	0.29 ± 0.42	0.20 ± 0.39

the prior mean function $m(\boldsymbol{\psi})$ equal to zero, the conditional distribution of $\mathbf{f}_*$ given the training set $\mathcal{S}_t$ is a multivariate normal distribution to which corresponds the posterior Gaussian process $\mathcal{GP}(m_t(\boldsymbol{\psi}), k_t(\boldsymbol{\psi}, \boldsymbol{\psi}'))$, in which

$$m_t\ (\boldsymbol{\psi}) = K\left(\boldsymbol{\psi}, \boldsymbol{\Psi}_t\right)\left[K\left(\boldsymbol{\Psi}_t, \boldsymbol{\Psi}_t\right) + \sigma_\delta^2 \mathbf{I}\right]^{-1} \mathbf{y}_t \tag{14}$$

$$k_t\ (\boldsymbol{\psi}, \boldsymbol{\psi}') = k\left(\boldsymbol{\psi}, \boldsymbol{\psi}'\right) - K\left(\boldsymbol{\psi}, \boldsymbol{\Psi}_t\right) \times \left[K\left(\boldsymbol{\Psi}_t, \boldsymbol{\Psi}_t\right) + \sigma_\delta^2 \mathbf{I}\right]^{-1} K(\boldsymbol{\psi}, \boldsymbol{\Psi}_t)^T \tag{15}$$

where $K(\boldsymbol{\psi}, \boldsymbol{\Psi}_t)$ is the row vector of the covariances between every training input and $\boldsymbol{\psi}$, $K(\boldsymbol{\Psi}_t, \boldsymbol{\Psi}_t)$ is the covariance matrix between all the training inputs, and $\mathbf{I}$ is the identity matrix. Both

TABLE XIV
GPR PERFORMANCE VARIATION WITH THE KERNEL FUNCTION; CASE STUDY C1, TURBINE A1, AND PCA-BASED PREDICTION, AS IN TABLE VI

	Kernel function		
PCA and GPR	Absolute exponential	Matérn, with $\nu = 5/2$	Squared exponential
LD_{ma} [dB] average ± std. dev.	1.39 ± 0.36	1.37 ± 0.37	1.40 ± 0.37
LD_{rms} [dB] average ± std. dev.	1.93 ± 0.43	1.92 ± 0.43	1.94 ± 0.43
ISD average ± std. dev.	0.18 ± 0.74	0.18 ± 0.50	0.26 ± 0.63

$K(\boldsymbol{\psi}, \boldsymbol{\Psi}_t)$ and $K(\boldsymbol{\Psi}_t, \boldsymbol{\Psi}_t)$ are built using the kernel function $k(\boldsymbol{\psi}, \boldsymbol{\psi}')$ [21]. Finally, to predict the output corresponding to a new input sample $\boldsymbol{\psi}$, the mean function of the posterior Gaussian process, $m_t(\boldsymbol{\psi})$, can be used. In the problem at hand, the input $\boldsymbol{\psi}$ is simply the wind speed w, and the output $\hat{y}$ is the value of the coefficient to be estimated.

Any positive-definite function can be used as kernel function: this flexibility is one of the benefits of GPR, allowing to reflect prior assumptions about the latent function [14]. The radial basis function (called also squared exponential), defined as

$$k\ (\boldsymbol{\psi}, \boldsymbol{\psi}') = \sigma_f^2\ \exp\left(-\frac{\|\boldsymbol{\psi} - \boldsymbol{\psi}'\|^2}{2\lambda^2}\right) = \sigma_f^2\ \exp\left(-\frac{\tau^2}{2\lambda^2}\right) \tag{16}$$

is a very popular choice to model very smooth and stationary functions. The signal variance σ_f^2 and the length scale λ can be tuned to govern the a priori correlation between points at an Euclidean distance $\tau = \|\boldsymbol{\psi} - \boldsymbol{\psi}'\|$. On the opposite side, the absolute exponential kernel

$$k\ (\boldsymbol{\psi}, \boldsymbol{\psi}') = \sigma_f^2\ \exp\left(-\frac{\tau}{\lambda}\right) \tag{17}$$

is adopted to model extremely unsmooth functions. A clear example of the difference between posterior Gaussian processes adopting these two kernels is given in [21]. The Matérn class of covariance functions (see [14] for the general definition) has an

TABLE XV
PERFORMANCE OF PREDICTION METHODS REPORTED IN TABLES II AND III FOR CASE STUDIES C1 AND C2, WITH THE ADDITION OF THE PCC VALUE (AVERAGE ± STD. DEV.)

	Case study C1, as in Table II			Case study C2, as in Table III		
	AAI	PCA	NMF	AAI	PCA	NMF
LD_{ma} [dB]	2.55 ± 1.00	1.38 ± 0.35	2.21 ± 0.63	2.18 ± 0.77	1.60 ± 0.42	2.30 ± 1.02
LD_{rms} [dB]	3.13 ± 0.98	1.93 ± 0.41	2.85 ± 0.62	2.74 ± 0.80	2.32 ± 0.48	2.91 ± 1.05
ISD	0.46 ± 0.41	0.18 ± 0.26	0.18 ± 0.08	0.34 ± 0.29	0.41 ± 1.43	0.47 ± 1.73
PCC	0.83 ± 0.09	0.89 ± 0.05	0.87 ± 0.05	0.85 ± 0.08	0.87 ± 0.06	0.82 ± 0.07

The green, orange and red boxes show the best, intermediate and worst prediction method, respectively.

additional parameter ν, which governs the smoothness of the resulting function. For $\nu = 1/2$, the Matérn covariance becomes identical to the absolute exponential kernel, whereas as $\nu \to \infty$, it converges to the radial basis kernel. The functions obtained with $\nu = 3/2$ and $\nu = 5/2$ are commonly adopted as useful compromises between the two extremes.

The values for the hyperparameters used in the GPR model (e.g., λ, σ_f^2, σ_δ^2) can be inferred from the training set data, maximizing the probability of the observed outputs, given the related inputs and the hyperparameters. This means to maximize the marginal log-likelihood function [14], [21], a problem commonly solved through a gradient-ascent based optimization scheme.

In this article, the Matérn kernel with $\nu = 5/2$ [14]

$$k\ (\boldsymbol{\psi}, \boldsymbol{\psi}') = \sigma_f^2 \left(1 + \frac{\sqrt{5}\tau}{\lambda} + \frac{5\tau^2}{3\lambda^2}\right) \exp\left(-\frac{\sqrt{5}\tau}{\lambda}\right) \quad (18)$$

was used, because it provided results that were slightly better than those obtained with the other kernels mentioned. This point is discussed in Section D of this appendix.

C. RFR Hyperparameters

The RFR results reported in this article were obtained with $\mathcal{E} = 30$ and $\ell = 8$, values that yielded good performance (relative to the results of the other techniques tested) in all three case studies. This does not mean that they are the best values for each case. For example, if the test for case study C1, the results of which are shown in Table VI, is repeated by using the 15 combinations obtained when $\mathcal{E}$ takes value in $\{10, 30, 50\}$ and ℓ in $\{4, 8, 12, 16, 20\}$, the best performance is obtained with $\mathcal{E} = 30$ and $\ell = 16$. Table XII shows the improvement in performance, which is now almost identical to that obtained with GPR (see Table VI). However, if the values $\mathcal{E} = 30$ and $\ell = 16$ are used instead of $\mathcal{E} = 30$ and $\ell = 8$ in the test for case study C2, the results of which are shown in Table VII, the results, instead of improving, slightly worsen (except ISD), as shown in Table XIII. This demonstrates that the values of the RFR hyperparameters that provide the best performance are different for different applications.

While GPR inherently optimizes hyperparameters during the training phase, by maximizing a likelihood function (see Section B of this appendix), a refined use of RFR would require adding specific hyperparameter optimization strategies, using the data available in the training phase.

D. GPR Kernels

The GPR results reported in this article were obtained using the class of Matérn kernels, with $\nu = 5/2$. Adopting the other kernels mentioned in Section A of this appendix, more or less smooth than the one used here, does not significantly change the results. For example, if the test for case study C1, the results of which are shown in Table VI, is repeated by adopting the squared exponential and absolute exponential kernels, the results worsen to a negligible extent. Table XIV reports the related values. This stability of the result with respect to the change of the kernel function is another strength of the GPR technique.

APPENDIX B

The Pearson correlation coefficient (PCC) is defined as the ratio between the covariance of two random variables and the product of their std. dev. It represents a normalized covariance measurement that, by construction, ranges in [–1, 1]. In statistics, it is used to measure the linear correlation between two sets of data: assuming a given spectrum at p frequency values to be a set of observations of a random variable, the similarity between the spectra $\mathbf{s}(w)$ and $\hat{\mathbf{s}}(w)$ can be assessed by their coefficient of correlation

$$\text{PCC}\ (w) = \frac{p\sum_{b=1}^{p} \mathbf{s}_b\hat{\mathbf{s}}_b - \sum_{b=1}^{p} \mathbf{s}_b \sum_{b=1}^{p} \hat{\mathbf{s}}_b}{\sqrt{\left[p\sum_{b=1}^{p} \mathbf{s}_b^2 - \left(\sum_{b=1}^{p} \mathbf{s}_b\right)^2\right]\left[p\sum_{b=1}^{p} \hat{\mathbf{s}}_b^2 - \left(\sum_{b=1}^{p} \hat{\mathbf{s}}_b\right)^2\right]}} \quad (19)$$

where the subscript b, $b \in [1, p]$, indicates the bth entry of the two vectors and the spectra wind dependence has been neglected for brevity. This similarity measurement (the closer PCC is to 1, the greater the similarity between the two spectra) is widely adopted in spectrometry, Raman spectroscopy, and multispectral remote sensing [22], [23], [24], [25].

In this appendix, PCC is compared to the three matching metrics introduced in Section II-F by repeating the tests described in Section IV-C, the results of which were reported in Tables II and III. Table XV adds the PCC values to the values of the three

metrics used in this article. It is easily seen that PCC is always aligned with LD_{ma} and LD_{rms}, leading to the same ranking of the compared prediction methods. PCC is only partially aligned with ISD, as are, on the other hand, LD_{ma} and LD_{rms}. Finally, according to PCC, performance worsens for PCA and NMF going from C1 to C2, while it improves for AAI. The same judgment is provided by each of the three metrics used in this article.

Since PCC does not seem to add significant information, only the three metrics that are most commonly adopted in assessing the similarity between acoustic spectra were used in this article.

ACKNOWLEDGMENT

The author would like to thank Tanja Pangerc, Peter. D. Theobald, Lian S. Wang, Stephen P. Robinson, and Paul A. Lepper for providing the data sets used in this study. The author would also like to thank Prof. Vittorio Murino for valuable advice, discussions, and revisions related to the topic and text of this manuscript.

REFERENCES

[1] T. A. Mooney, M. H. Andersson, and J. Stanley, "Acoustic impacts of offshore wind energy on fishery resources," *Oceanography*, vol. 33, no. 4, pp. 82–95, 2020.

[2] U. Stöber and F. Thomsen, "How could operational underwater sound from future offshore wind turbines impact marine life?," *J. Acoust. Soc. Amer.*, vol. 149, no. 3, pp. 1791–1795, 2021.

[3] J. C. Svendsen, B. Ibanez-Erquiaga, E. Savina, and T. Wilms, "Effects of operational off-shore wind farms on fishes and fisheries," DTU Aqua, Kongens Lyngby Denmark, DTU Aqua Rep. 411-2022, 2022.

[4] Y. T. Lin, A. E. Newhall, J. H. Miller, G. R. Potty, and K. J. Vigness-Raposa, "A three-dimensional underwater sound propagation model for offshore wind farm noise prediction," *J. Acoust. Soc. Amer.*, vol. 145, no. 5, pp. EL335–EL340, 2019.

[5] J. Tougaard, L. Hermannsen, and P. T. Madsen, "How loud is the underwater noise from operating offshore wind turbines?," *J. Acoust. Soc. Amer.*, vol. 148, no. 5, pp. 2885–2893, 2020.

[6] K. Betke and M. A. Bellmann, "Operational underwater noise from offshore wind farms," in *The Effects of Noise on Aquatic Life*, A. N. Popper, J. Sisneros, A. Hawkins, and F. Thomsen, Eds., Berlin, Germany: Springer, 2023.

[7] F. Thomsen, U. Stöber, and J. Sarnocińska-Kot, "Hearing impact on marine mammals due to underwater sound from future wind farms," in *The Effects of Noise on Aquatic Life*, A. N. Popper, J. Sisneros, A. Hawkins, and F. Thomsen, Eds., Berlin, Germany: Springer, 2023.

[8] T. Pangerc, P. D. Theobald, L. S. Wang, S. P. Robinson, and P. A. Lepper, "Measurement and characterisation of radiated underwater sound from a 3.6 MW monopile wind turbine," *J. Acoust. Soc. Amer.*, vol. 140, no. 4, pp. 2913–2922, 2016.

[9] R. Wang, X. Xu, Z. Zou, L. Huang, and Y. Tao, "Dominant frequency extraction for operational underwater sound of offshore wind turbines using adaptive stochastic resonance," *J. Mar. Sci. Eng.*, vol. 10, no. 10, 2022, Art. no. 1517.

[10] T. Soukissian, F. Karathanasi, and P. Axaopoulos, "Satellite-based offshore wind resource assessment in the Mediterranean Sea," *IEEE J. Ocean. Eng.*, vol. 42, no. 1, pp. 73–86, Jan. 2017.

[11] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. Berlin, Germany: Springer, 2009.

[12] J. O. Rawlings, S. G. Pantula, and D. A. Dickey, Eds., *Applied Regression Analysis: A Research Tool*. Berlin, Germany: Springer, 1998.

[13] C. De Boor, *A Practical Guide to Splines (Revised Edition)*. Berlin, Germany: Springer, 2001.

[14] C. E. Rasmussen and C. K. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA, USA: MIT Press, 2006.

[15] D. Lee and H. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788–791, 1999.

[16] D. N. Hào and D. Lesnic, "Heuristic regularization methods for numerical differentiation," *Comput. Math. Appl.*, vol. 63, no. 4, pp. 816–826, 2012.

[17] L. R. Rabiner and B.-H. Juang, *Fundamentals of Speech Recognition*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.

[18] V. V. Savchenko, "Itakura–Saito divergence as an element of the information theory of speech perception," *J. Commun. Technol. Electron.*, vol. 64, pp. 590–596, 2019.

[19] S. L. Marple Jr., *Digital Spectral Analysis With Applications*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1987.

[20] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[21] E. Schulz, M. Speekenbrink, and A. Krause, "A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions," *J. Math. Psychol.*, vol. 85, pp. 1–16, 2018.

[22] F. M. Howari, "Comparison of spectral matching algorithms for identifying natural salt crusts," *J. Appl. Spectrosc.*, vol. 70, no. 5, pp. 782–787, 2003.

[23] F. Van der Meer, "The effectiveness of spectral similarity measures for the analysis of hyperspectral imagery," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 8, no. 1, pp. 3–17, 2006.

[24] P. R. Griffiths and L. Shao, "Self-weighted correlation coefficients and their application to measure spectral similarity," *Appl. Spectrosc.*, vol. 63, no. 8, pp. 916–919, 2009.

[25] A. Z. Samuel et al., "On selecting a suitable spectral matching method for automated analytical applications of Raman spectroscopy," *ACS Omega*, vol. 6, no. 3, pp. 2060–2065, 2021.

Andrea Trucco (Senior Member, IEEE) was born in Genoa, Italy, on February 10, 1970. He received the M.Sc. and Ph.D. degrees in electronic and computer engineering from the University of Genoa, Genoa, in 1994 and 1998, respectively.

Since 1999, he has been with the University of Genoa, where he became a Professor in telecommunications in 2016 and was the Vice-Rector for International Relations from 2015 to 2020 and the Vice-Director of the Department of Biophysical and Electronic Engineering from 2008 to 2012. He was also a Senior Researcher with the Italian Institute of Technology, Genoa, from 2011 to 2015. He was a principal investigator in many research projects, funded by the European Union, national agencies, and industrial companies. He is the author of more than 200 scientific papers (95 in international journals) and inventor of seven international patents. His research interests include acoustics, signal processing, pattern recognition, sonar systems, medical ultrasound, and fault diagnosis.

Prof. Trucco won the student paper competitions organized by the 9th International Symposium on Unmanned Untethered Submersible Technology in 1995 and the MTS/IEEE Oceans Conference in 1997. He is an Associate Editor for IEEE JOURNAL OF OCEANIC ENGINEERING and IEEE TRANSACTIONS ON ULTRASONICS, FERROELECTRICS, AND FREQUENCY CONTROL.

Open Access funding provided by 'Università degli Studi di Genova' within the CRUI CARE Agreement