energies

MDPI

Article

Normal Behaviour Models for Wind Turbine Vibrations: Comparison of Neural Networks and a Stochastic Approach

Pedro G. Lind [1,*], Luis Vera-Tudela [2], Matthias Wächter [2], Martin Kühn [2] and Joachim Peinke [2]

1 Institut für Physik, Universität Osnabrück, Barbarastrasse 7, 49076 Osnabrück, Germany
2 ForWind—Center for Wind Energy Research, Institute of Physics, Carl von Ossietzky University of Oldenburg, Küpkersweg 70, 26129 Oldenburg, Germany; luis.vera.tudela@forwind.de (L.V.-T.); matthias.waechter@uni-oldenburg.de (M.W.); Martin.Kuehn@forwind.de (M.K.); joachim.peinke@uni-oldenburg.de (J.P.)
* Correspondence: pedro.g.lind@gmail.com; Tel.: +49-(0)541-969-2586; Fax: +49-(0)541-969-3472

Received: 29 October 2017; Accepted: 17 November 2017; Published: 23 November 2017

Abstract: To monitor wind turbine vibrations, normal behaviour models are built to predict tower top accelerations and drive-train vibrations. Signal deviations from model prediction are labelled as anomalies and are further investigated. In this paper we assess a stochastic approach to reconstruct the 1 Hz tower top acceleration signal, which was measured in a wind turbine located at the wind farm Alpha Ventus in the German North Sea. We compare the resulting data reconstruction with that of a model based on a neural network, which has been previously reported as a data-mining algorithm suitable for reconstructing this signal. Our results present evidence that the stochastic approach outperforms the neural network in the high frequency domain (1 Hz). Although neural network retrieves accurate step-forward predictions, with low mean square errors, the stochastic approach predictions better preserve the statistics and the frequency components of the original signal, retaining high accuracy levels. The implementation of our stochastic approach is available as open source code and can easily be adapted for other situations involving stochastic data reconstruction. Based on our findings we argue that such an approach could be implemented in signal reconstruction for monitoring purposes or for abnormal behaviour detection.

Keywords: wind turbine; tower acceleration; condition monitoring; signal reconstruction; neural networks; stochastic modelling

1. Introduction

As the number of wind turbines installed worldwide increases, the reduction of operational and maintenance cost by enhanced surveillance gains on importance [1]. Recent reviews on condition monitoring of wind turbines [2–4], recognize the importance that a reliability based maintenance system has on reducing the cost of energy. Although it is possible to monitor wind turbine functioning with 10-min average SCADA data, part of the information is lost when data is averaged; a mixed system, combining SCADA data and conventional high frequency sensors is expected in the future [3].

Zaher et al. [5] described the lack of failure records as the main challenge for early failure detection, which emphasizes the importance of normal behaviour modelling. Such models are built empirically, i.e., using only measured signals, and refer to the expected value the signal should have based on its own historical records. An anomaly in the signal is detected when the error on its prediction, made by the normal behaviour model, increases over several subsequent values.

Vibrations in wind turbines reduce their performance and their expected lifetime; thus, strategies developed for large structures, as buildings and brigdes, are applied and optimised to control their

vibration. Dampers are incorporated to wind turbines as part of a passive, active or semi-active strategy [6]. Also, their structural behaviour is recorded with sensors, which serve to assess its normal functioning, which is a challenge on its own due to the non-stationary nature of the excitation. Metamodels are foreseen to model short-term variations, which are governed by the rotation of components, as well as long-term variations, which are governed by the changing nature of excitation loads [7].

Previous investigations have built normal behaviour models with data-mining algorithms on SCADA data. Most of them are based on 10-min data [8]. In Ref. [9], the authors presented their application to detect anomalies in wind turbine gearboxes. In Ref. [5] the authors proposed a framework for their combination into a single user interface to facilitate their use, while Ref. [10] shows that a normal behaviour model based on neural networks (NN) outperforms a regression based one. However, some scientists also evaluate data with higher sampling frequency, e.g., in [11] the authors analysed the performance of various NN algorithms to detect bearing faults on wind turbines. In Refs. [12,13], they evaluated wind turbine vibrations in time domain with 10-s sampling signals from drive-train vibrations and tower top accelerations.

Furthermore, the capability of NN as means to monitor normal behaviour models for condition monitoring of wind turbines can be further enhanced updating preprocessing and post-processing steps, and so improving their training and to reducing their false alarms [14].

In contrast to existing methods, signals can also be reconstructed based on a stochastic approach, which has not been assessed so far for the creation of normal behaviour models. In a previous work [15], we demonstrated its value to estimate fatigue loads for wind turbines based on reconstructed signals. Though this stochastic approach was already combined with NN models [16] for improving forecast of air quality, there is no systematic investigation for evaluating whether the stochastic approach alone can be used as an alternative to other standard monitoring procedures, namely using NN. We focus specifically on assessing the stochastic approach as basis for normal behaviour models. We evaluate the performance of such model to monitor vibrations on wind turbines and compare it versus one based on a NN, which was found the most suitable NN model to reconstruct the tower top acceleration [13]. Our goal is to present the stochastic approach in terms than can be understood by a larger audience, and thus the comparison is not intended to find the optimum model, by building and comparing several regression models, as it was carried out in [11], but to build upon their findings and communicate accordingly. We focus our assessment on the similarities and differences between both approaches, and then discuss our findings with respect to the limitations imposed by the specific case evaluated.

We start in Section 2 by describing in detail the data used in this investigation, obtained from one wind turbine in an offshore wind farm. In Section 3 both the NN model and our stochastic approach are introduced; then, the results obtained from them are properly presented. In Section 4 we compare the results of both approaches when reconstructing one month of the tower acceleration measured at one turbine. Section 5 concludes the paper.

2. Data Description

Wind turbine vibrations can easily be assessed monitoring two standard operational signals: drive-train and tower top acceleration [13]. In this paper, we limited the analysis to the second one to assess the reconstruction of signals with both the neural network model and the stochastic approach.

The data analysed in this paper correspond to measurements carried out in one 5 MW Senvion wind turbine AV-04 of the offshore wind farm Alpha Ventus [17]. Alpha Ventus is the first offshore wind farm in Germany, located at Borkum West in the North Sea, 54.3° N–6.5° W. See Figure 1.

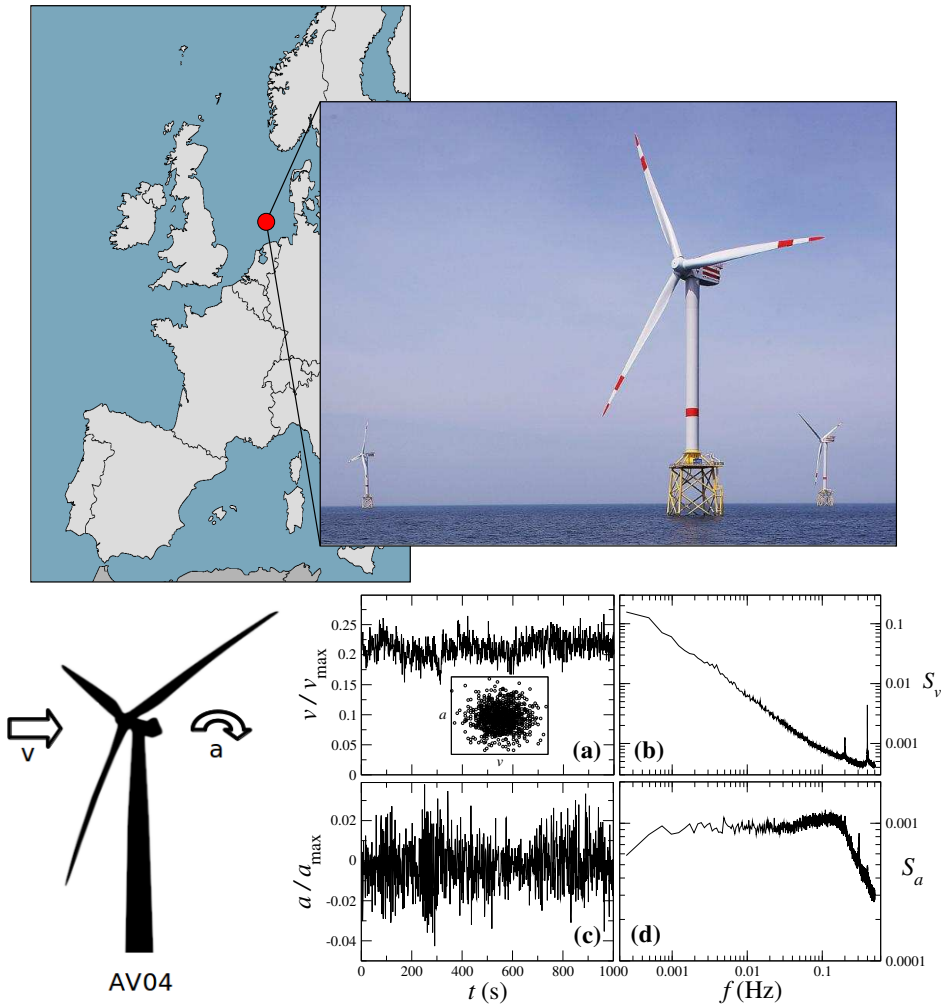

Figure 1. Location of the Alpha Ventus wind farm at Borkum West in the North Sea (54.3° N–6.5° W) together with a picture of turbine AV04. Examples of (**a**,**b**) wind speed and (**c**,**d**) tower acceleration signals in Alpha Ventus wind farm. The left ones ilustrate the time-domain of both quantities while in the right the frequency domain is shown for the full datasets of October 2014. Scatter plot $v \times a$ is shown in the inset of (**a**). All data were normalized to fulfill all confidentiality protocols, namely normalization to maximal values (see text). Both wind velocity and tower acceleration are normalized to maximal observed values, v_{max} and a_{max} respectively.

The turbine is subjected most of the time to undisturbed inflow due to its position in the first row of turbines, located at the western past of the wind farm. The support structure of the turbine combines a rather stiff and hydrodynamically transparent jacket structure with a tubular tower on top. In such a design the tower top acceleration signal is dominated by the dynamic response on aerodynamic excitation mainly at the first bending natural frequency of the system and at multiples of the blade passing frequency. Wave excitation plays only a minor role. The blade passing frequency is 0.175 Hz, while the first eigenfrequency of the tower is 0.34 Hz [18].

Data include all wind speed and tower top acceleration values recorded in October and November 2014, regardless of the operational and wind flow conditions. Therefore, 29 days were available in October (2,555,805 values) and 21 days in November (1,859,179 values). We only removed incorrect measurements from the data collected, either because they were identified by the recording system with a flag ('99999' in our case) or because their values were out of reasonable physical values, namely $x > \bar{x} \pm 5\sigma$. Wind speed and tower acceleration values were available with sampling ratios of 1 Hz and 50 Hz respectively, thus the latter one was down-sampled to 1 Hz in order to complete the analysis.

From measurements, 1434 and 992 records were flagged as '99999' (equivalent to less than 24 min) while 1 and 45 (less than a minute) were found outside reasonable physical values and removed from datasets representing the whole months for October and November respectively.

Wind speeds were measured by a cup anemometer included in the meteorological station located on top of the nacelle. Tower top longitudinal accelerations were recorded by an accelerometer positioned at the bottom of the nacelle, near its connection to the upper part of the tower. Figure 1 illustrates the characteristics of the data in time and frequency domain. All measurements are normalized to maximal values.

The cup anemometer is suitable for 1 Hz sampling, similar to previous studies [15,19,20] to analyse fatigue loads, dynamic power curves and damages on wind turbines. It turned out that this sampling frequency is indeed sufficient to resolve the response dynamics of turbine quantities with respect to the wind fluctuations.

Since we aimed to construct a model for prediction and not to explain or find causality, we did not perform any correction on the signals reconstructed. Although models that explain are assumed to have good predictive characterstics, the first one is not a requirement to have the second one, as discussed in Ref. [21].

To protect sensitive manufacturer information we normalized the values. Wind speeds recorded included the whole operational range of the wind turbine. Furthermore, we divided measurements in two sets according to different purposes: data collected in October were used to estimate the parameters in each model, while data collected in November were used to test each model's performance.

3. Methods and Results

Each approach to construct a normal behaviour model is introduced in a concise manner. Where convenient, we provide references to literature with extensive details about neural networks and the stochastic approach. At the end of the section, we describe the performance metrics utilised for the assessment.

In [13], the authors evaluated a pool of potential input variables to predict tower top acceleration signal with a wrapper algorithm [22], which includes the predictor model to search for the variables that reduce prediction error (*a posteriori* approach). They found wind speed, tower acceleration and wind direction relevant. Since our evaluation focused on the approaches themselves, we limited the pool of potential input variables to a single external signal (wind speed) and previous values of the tower top acceleration to have the same number of inputs for each approach.

3.1. Neural Networks: A Deterministic Approach

Neural networks transform input variables with a linear coefficient and a non-linear function. This is repeated, in a chain, over a variable number of layers, where each output layer becomes the input of the next one. The output of the last layer is the final prediction made for the target value. In this analysis, we utilised a non-linear auto-regressive with exogenous inputs (NARX) neural network, which is a recurrent network, i.e., it includes one or more feedback loops. Thus, the approach followed to estimate the predicted value of the tower top acceleration, given by:

$$\hat{a}(t) = f[a(t-\Delta t), \ldots, a(t-n_a\Delta t), v(t-\Delta t), \ldots, v(t-n_v\Delta t)] + e(t), \quad (1)$$

involves the use of its n_a previous values $(a(t-\Delta t), \cdots, a(t-n_a\Delta t))$ and those of the n_v previous wind speeds $(v(t-\Delta t), \cdots, v(t-n_v\Delta t))$ through a smooth unknown function $f()$. An extra random component $e(t)$ represents a random error, which has a zero mean and is independent of v and a [23]. The time-dependent predicting function $\hat{a}(t)$ can be taken as an evolution equation.

To create the baseline normal behaviour model we used a NN to approximate $f()$, which is a recursive non-linear transformation of its inputs. In Ref. [13], the authors demonstrated that NNs were the most suitable algorithm to create a normal behaviour model for the tower top acceleration of a

wind turbine when compared with neural network ensemble, boosting regression trees, support vector machine, random forest with regression, standard classification and regression tree and K-nearest neighbour. In our investigation, we used a NN with two layers (two transformations). The first one (hidden layer) consisted of n_h neurons, which apply non-linear transformations via a sigmoid function; the second one (output layer) was formed by a single neuron, which performed an extra linear transformation, with given weights w_i.

To optimize the neural network we can also include a finite number of feedback loops for previous values of the tower top accelerations (a, n_a) and wind speeds (v, n_v), thus we have three parameters to determine (n_h, n_v, n_a).

First, we divided data from October 2014 using the hold out method [22] in sub-sets for model training (70%), validation (20%) and test (10%). Hold out was selected also at this stage [16] because of the large number of points available in October and the data hold out from November. This was meant to be in agreement with our intention to keep the inter model variations as small as possible. Then, we selected the number of neurons in the hidden layer (n_h), the number n_v of input delays for the wind speed v and the number n_a of output delays for the tower top acceleration a with a wrapper algorithm [24], which searched for optimum prediction in the ranges $n_h, n_a, n_v \in \{1, \cdots, 50\}$.

The optimization of the unknown function f was set to minimize mean squared error:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2, \tag{2}$$

using the Levenberg-Marquardt algorithm. As a result, the architecture of the NARX network consisted on 25 neurons in the hidden layer ($n_h = 25$), three input delays ($n_v = 3$) and one output delay ($n_a = 1$).

Figure 2 depicts the development and utilization of the NN model investigated. From left to right, first we used the three previous wind speeds and tower top accelerations collected in October 2014 to define the constitutive components of the model, which is named a neural net black box in Figure 2. Once the evolution equation $\hat{a}(t)$ was defined, we used wind speeds and the previous acceleration value collected in November 2014 to predict the tower top acceleration signal, as well as to evaluate the performance of the model.

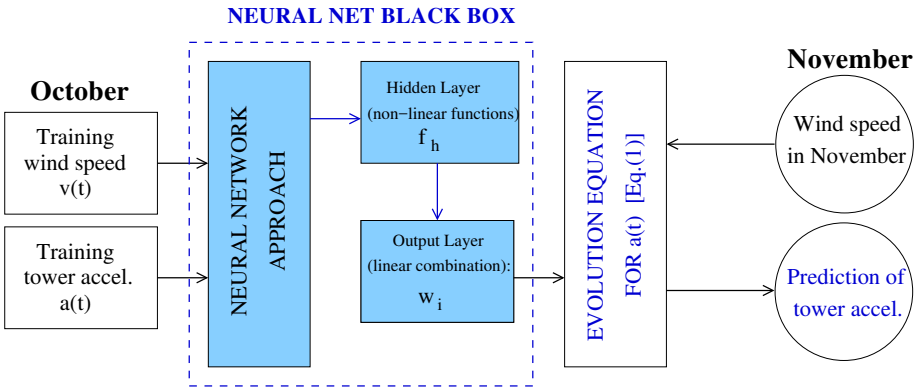

Figure 2. Schematic illustration of the neural network (NN) approach, indicating the development and test of neural networks for signal reconstruction. Here the particular case of tower acceleration reconstruction is given.

3.2. Stochastic Approach: The Langevin Model

The Langevin model is a framework developed from the pioneering work in Refs. [25,26], which consists of a direct method for extracting the evolution equation of stochastic series of measurements. Several applications were proposed and developed, e.g., in turbulence modelling, in medical EEG monitoring and in stock markets. See Ref. [27] for a review. In the context of wind

energy, this framework has shown the ability for predicting power curves of single wind turbines as well as of equivalent power curves for entire wind farms, and also to properly reproduce the increment statistics of power and torque in single wind turbines from wind speed measurements [19].

Figure 3 depicts the development and utilization of the stochastic approach investigated. Our stochastic framework, instead of computing non-linear functions and weights, which optimize a set of output compared to input data, retrieves two single functions of the variables involved. One of such functions, below symbolized by $D^{(1)}$, governs the deterministic contribution for the time variation of the output variable, while the other function, $D^{(2)}$, accounts for the stochastic fluctuations that include all the non-observable degrees of freedom present in the system. See the illustration in Figure 4.

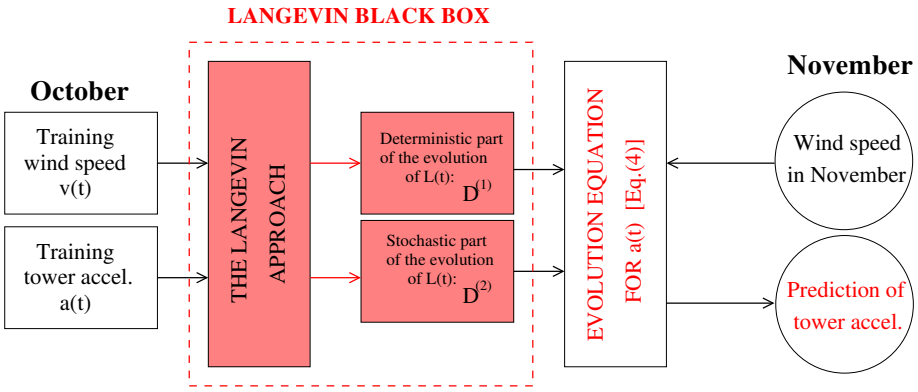

Figure 3. Schematic illustration of the Langevin (stochastic) approach. Here the particular case of tower acceleration reconstruction at Alpha Ventus is given.

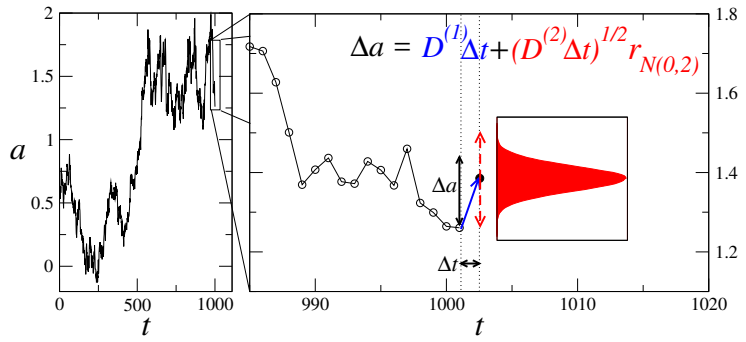

Figure 4. Illustration of the stochastic approach, while integrating the evolution Equation (4), composed by its two contributions, the deterministic contribution $D^{(1)}$ which governes the tendency of the evolving observable (blue), and the stochastic fluctuations accounted by $D^{(2)}$ (red) added to the deterministic part.

Having wind speeds $v(t)$ and tower accelerations $a(t)$, the following evolution equation for the latter quantity is used:

$$\frac{da}{dt} = D^{(1)}(a,v) + \sqrt{D^{(2)}(a,v)}\Gamma_t, \tag{3}$$

with Γ_t representing a Gaussian δ-correlated white noise, i.e., $\langle \Gamma_t \rangle = 0$ and $\langle \Gamma_t \Gamma_{t'} \rangle = 2\delta(t-t')$.

The reconstruction of the tower acceleration $a(t)$ follows therefore directly from the stochastic integration of Equation (3) which yields:

$$a(t+\Delta t) = a(t) + D^{(1)}(a,v)\Delta t + \sqrt{D^{(2)}(a,v)\Delta t}\; r_{N(0,2)}, \tag{4}$$

where $r_{N(0,2)}$ is a random number from a Gaussian distribution with zero mean and variance equal to two. Details are illustrated in Figure 4 and can be found in Ref. [27].

Function $D^{(1)}$ is also called drift function, while function $D^{(2)}$ is typically named as the diffusion function. Since the drift and diffusion functions have a physical interpretation, one could apply the model in Equation (3) to a particular system and define *ad hoc* the functional shape of both functions from physical reasoning. Next we explain how to compute these both functions.

The derivation is performed through the computation of the corresponding first and second conditional moments [15], respectively:

$$M^{(1)}_{v^*}(a,\tau) = \langle a(t+\tau) - a(t) \rangle \,|_{a(t)=a, v(t)=v^*}, \tag{5a}$$

$$M^{(2)}_{v^*}(a,\tau) = \left\langle (a(t+\tau) - a(t))^2 \right\rangle |_{a(t)=a, v(t)=v^*}. \tag{5b}$$

where $\langle \cdot \rangle|_{X(t)=x}$ indicates the average over the full time series, whenever $X(t)$ takes the value x, defined within an interval.

Figure 5a shows the first conditional moment for different values of the tower acceleration, namely for $a = 0.0075 a_{max}$ (bin 1), $a = -0.0025 a_{max}$ (bin 2), $a = 0.01 a_{max}$ (bin 3), computed as given in Equation (5). Typically, for the lowest values of τ one observes a linear dependence of the conditional moments with τ. Since it can be shown that drift ($D^{(1)}$) and diffusion functions ($D^{(2)}$) in Equation (3) are, apart from a multiplicative constant ($1/k!$), the derivative with respect to the time-gap τ of the first and second conditional moments respectively, we can define both functions directly from the conditional moments, namely:

$$D^{(k)}(a,v) = \lim_{\tau \to 0} \frac{1}{k!} \frac{M^{(k)}_v(a,\tau)}{\tau}. \tag{6}$$

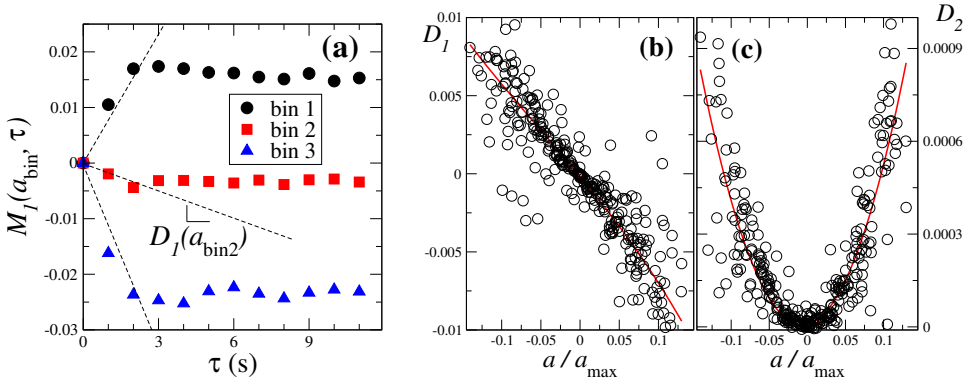

Figure 5. Illustration of (**a**) the first conditional moment for three different bin values of the tower acceleration as defined in Equation (5). Here v^* is given by the average velocity found during October. Numerical result for (**b**) the drift $D^{(1)}(a,v)$ and (**c**) the diffusion $D^{(2)}(a,v)$ in the Langevin equation given by Equation (6), plotted as function of a alone, i.e., they are projected at the a-axis to emphasize the linear and quadratic dependency of $D^{(1)}$ and $D^{(2)}$ respectively for the largest range of acceleration values. In red a polynomial fit of both functions is shown. We use bins for the velocity with a width of 0.017 in units of maximal velocity while acceleration was taken in bins of width 0.017 of maximal tower acceleration.

In other words, by taking the slope of the linear regression for each conditional moment one arrives to the corresponding value of function $D^{(1)}$ and $D^{(2)}$. Indeed, as sketched in Figure 5a, within a sufficiently low range of τ values, here three time-steps of the set of measurements, the conditional moments $M^{(1)}$ and $M^{(2)}$ depend linearly on τ. Figure 5b,c show both drift and diffusion functions, ($D^{(1)}$) and ($D^{(2)}$) respectively, for a range of a-values at different velocities. Notice that while the diffusion function shows a quadratic dependence on the tower acceleration, the drift function exhibits a cubic dependence with a dominant linear term having one single fixed point at $a^* = 0$ which corresponds to the equilibrium position of tower vibrations.

More precisely, a polynomial fit of $D^{(1)}$ yields:

$$D^{(1)}(a) = \alpha - ka\left(1 + \beta a + \gamma a^2\right) \tag{7}$$

with $\alpha = -1.75 \times 10^{-4}$ us^{-1}, u taken in units of a_{max}, with $k = 0.061$ s^{-1}, $\beta = 0.67$ u^{-1} and $\gamma = 4.9$ u^{-2}. Since the values of tower acceleration in units of u are typically of the order of 10^{-1} the dominant term in the cubic fit is clearly the linear one, yielding:

$$D^{(1)}(a) \sim -ka. \tag{8}$$

In other words, the "force" $D^{(1)}$ governing the drift of tower acceleration can be interpreted as a spring force with spring constant k around a stationary state with no acceleration that corresponds obviously to the vertical position of the tower. This interpretation of the stochastic model will be of importance below, where we discuss the possibility to use the stochastic modelling for monitoring tower vibrations and eventual structural defects of the tower. The fit of diffusion yields a quadratic polynomial $D^{(2)}(a) = d_0 + d_1 a + d_2 a^2$ with $d_0 = 7 \times 10^{-6}$, $d_1 = 5.4 \times 10^{-4}$ and $d_2 = 4.6 \times 10^{-2}$.

It is important to stress that our framework is based on the assumption that the observable follows a Markov process, which means nothing else than that the noise Γ_t is δ-correlated as mentioned above. Though, one advantage of this Langevin data reconstruction is that it also works in some cases where the Markov test fails [28]. The full implementation in one- and two-dimensions of this approach is already public available [29]. Improvements on the noise term are beyond this paper, but were already addressed previously [30,31].

3.3. Performance Evaluation

To compare both approaches, we consider the distribution of all values predicted from each model for November 2015, in particular their four first statistical moments, namely the mean, the standard deviation, the skewness and the kurtosis. Notice that, while mean and standard deviation are sufficient for characterizing the distribution of Gaussian processes, in general, higher-order moments should be checked for ascertaining if the process is non-Gaussian or not. The corresponding results are shown in Figure 6 and Table 1.

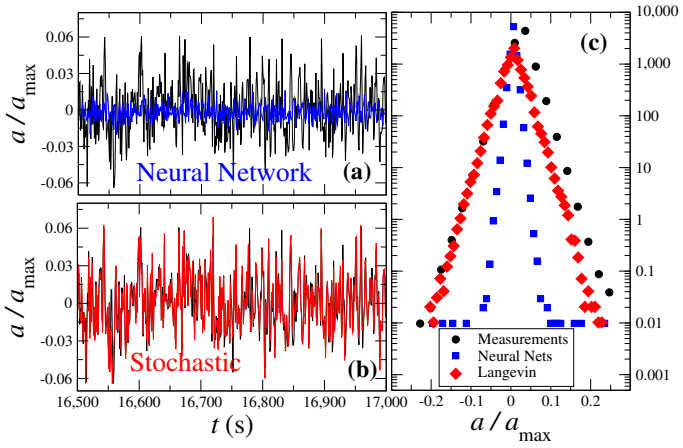

Figure 6. Sample of data series of tower acceleration from measurements at AV04 in Alpha Ventus and the acceleration reconstruction models using (**a**) NARX neural networks and (**b**) the Langevin model. In both cases, one also plots (black lines) the corresponding series of measurements for better comparison. The corresponding value distribution of these series is displayed in (**c**) with symbols for the full month of November 2014.

Table 1. First four statistical moments of the value distributions shown in Figure 6d for the normalised measurements and the reconstructed signals with each one of both models.

Signal	Mean	Std. Dev.	Skewness	Kurtosis
Measurements	-2.07×10^{-3}	24.7×10^{-3}	11.6×10^{-3}	5.07
NN	-2.15×10^{-3}	7.39×10^{-3}	-30.1×10^{-3}	6.31
Langevin model	-1.71×10^{-3}	25.6×10^{-3}	-3.59×10^{-3}	2.37

Additionally, to evaluate the prediction of both models we use four different metrics, namely the mean absolute error:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i|, \tag{9}$$

its standard deviation:

$$SDofAE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(|\hat{y}_i - y_i| - \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i|\right)^2}, \tag{10}$$

the mean square error (already presented in Section 3, Equation (2)) and its standard deviation:

$$SDofSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[(\hat{y}_i - y_i)^2 - \sum_{i=1}^{n}(\hat{y}_i - y_i)^2\right]^2}, \tag{11}$$

where $\hat{y}_i$ denotes the ith measured value of the property y, and y_i the corresponding modelled value, either with NN or with the Langevin model.

Notice that the standard coefficient of determination defined as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}, \tag{12}$$

with $\bar{y} = (1/n)\sum_{i=1}^{n}\hat{y}_i$, is related to MSE by:

$$MSE = \frac{1 - R^2}{n}\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2. \tag{13}$$

References [13,32] reported these metrics to select the best data-mining technique and to derive models for the normal behaviour of wind turbine vibration and to detect faults in wind turbine gearboxes. Moreover, they have been commonly used to evaluate the accuracy of models utilised for regression analysis in wind energy applications [11–13,32]. The correspondig results comparing the NN and the stochastic approach are summarized in Table 2.

Table 2. Performance of both models, using the metrics defined in Equations (2), (9)–(11), namely the mean of the absolute error, Equation (9), its standard deviation, Equation (10), the mean square error, Equation (2) and its standard deviation, Equation (11).

Signal	MAE	SDofAE	MSE	SDofSE
NN	0.031	0.032	2.0×10^{-3}	65.6×10^{-3}
Langevin model	0.027	0.031	1.6×10^{-3}	0.7×10^{-3}

Finally, we also analyse the temporal correlations of the data. Therefore we use two methods, namely we calculate the power spectra and analyse the statistics of the temporal increments:

$$\Delta a(t,\tau) = a(t+\tau) - a(t), \tag{14}$$

where $\tau = n\Delta t$ is an integer number n of consecutive time-steps Δt. Note that the second moment of the increments Δa, given by $\langle(\Delta a)^2\rangle$, corresponds to the power spectrum. The power spectrum of the reconstructed signals are plotted in Figure 7, showing how the frequency components are reconstructed. The statistics of the increments is given in Figure 8, from which we will ascertain how good the model retrieves the evolution of the process throughout the succession of measurements [27].

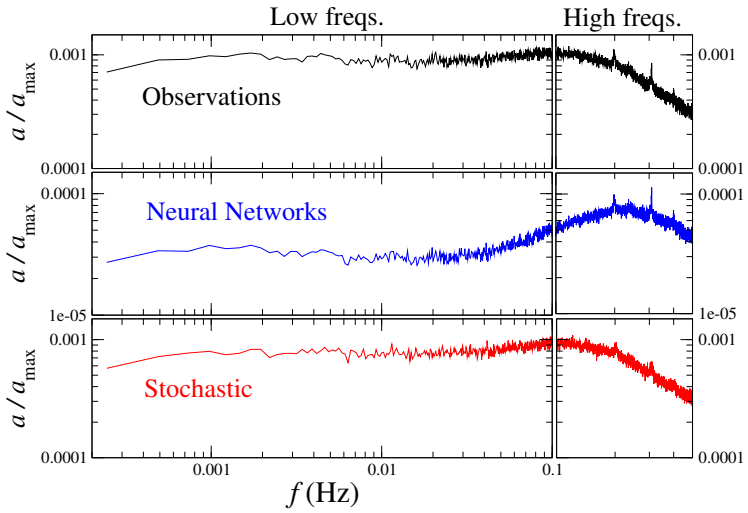

Figure 7. Spectrum of original signal (**top**) and the corresponding reconstructed signals, one using the neural network (**middle**) and the other using our proposed stochastic approach (**bottom**). Notice that the high frequency domain is plotted in a linear frequency scale, while the low frequency domain is plotted in the logarithmic scale.

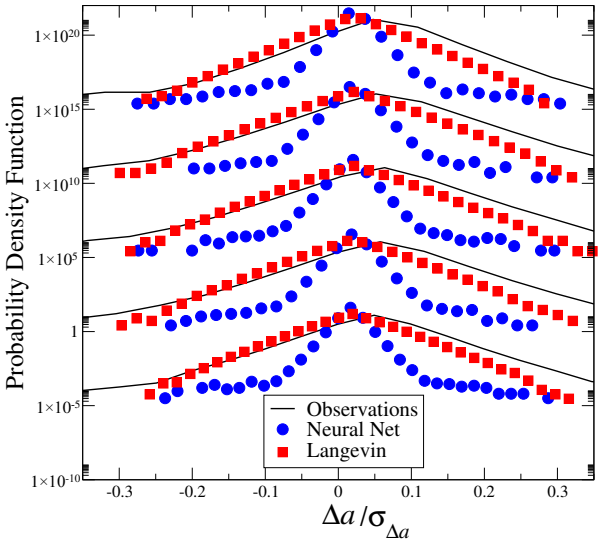

Figure 8. Two-point statistics of the tower acceleration (lines) and the corresponding reconstructed signals, i.e., value distributions of $\Delta a(t) = a(t+\tau) - a(t)$. From top to bottom one has $\tau = 1, 2, 4, 8$ and 16 s. The vertical shift of the distribution is for better visualization.

4. Discussion

4.1. Comparing Both Approaches

To evaluate the suitability of the stochastic approach, described above, for modelling normal behaviour of wind turbine vibrations, we compare it with the neural network [13].

Figure 6a,b shows normalized samples of the original tower top acceleration and its two reconstructions. Already in these plots it is clear the stronger resemblance between the reconstructed data from the stochastic approach and the set of tower acceleration measurements. While the NN reconstruction keeps the average value of the tower oscillations, their range of amplitudes is not as well reproduced as with the stochastic approach.

This feature is more evident in the plot of Figure 6c: both the stochastic and the NN models retrieve the approximate distribution of values of the original signal (symbols), but the full range of the tower acceleration values observed in measured distribution (bullets) is reproduced by the distribution obtained from the Langevin model (diamonds) but not from the NN model (squares).

Furthermore, the empirical distribution is clearly non-Gaussian, showing exponential tails. Table 1 compares the first four moments of all three distributions in Figure 6c. While NN reproduces the mean more accurately, the standard deviation is much better modelled with the stochastic approach. In other words, while the average behaviour is better modelled with NN, for properly reproduced the amplitude of tower vibrations the stochastic approach is better suited.

Curiously however, while the (weak) skewness is also better approximated by the stochastic approach, the value of the kurtosis is better reproduced by the NN model. Additionally, the response obtained with NN had a smaller amplitude, as the optimization of its parameters reduce its cost function, the mean square error, thus it tends to center its prediction to improve its accuracy. These statistics are important to better describe the shape of distributions and thus, indicate their symmetry and the weight of the tails, respectively. Our distribution is fairly symmetrical but with tails that differ from a normal distribution.

The performance of both models was also assessed using the metrics introduced above, Equations (9)–(11) and results are given in Table 2: one clearly sees a tendency for lower values of *MAE*, *SDofAE*, *MSE* and *SDofSE*. For such metrics, one concludes that the stochastic approach results have comparable accuracy or are even more accurate than the results from NN models, in what concerns the predictions for the tower top acceleration.

As for the temporal correlations of the signals, the results plotted in Figure 7 show the spectral behaviour of the two models compared with the measurements of the tower acceleration. While NN better reproduces the periodic modes (peaks in the spectrum) present in the measurements, it is worse for the low-frequency range and for the overall amplitudes of the spectrum. Notice the different vertical scale used for plotting the NN results. The stochastic approach smoothes out the periodic modes, but reproduces the overall shape of the empirical power spectrum.

The metrics chosen indicate that the stochastic approach is a good alternative to model the normal behaviour of a signal. However, for certain situations other approaches dealing with the identification of normal and abnormal behaviour may also be suited. For example, were we in posession of abnormalities in our data set, we could have also used a classification algorithm to mark records into either one of two classes (normal, abnormal). In such a case, care must been taken to account for the imbalance of instances in one of the classes [22], and as a consequence, we would have had to select other metrics, such as the F-measure as recommended by the investigation of fault diagnosis in gearboxes [33].

A better reconstruction of two-point statistics is also obtained through the stochastic approach, as shown in Figure 8. It is clear that the stochastic approach better reconstructs the differences Δa between consecutive values in the tower acceleration series, only with some small deviations for the largest differences. This shows furthermore that the assumption of Markovian process with Langevin noise is considerably reliable for these data sets.

4.2. Stochastic Modelling Applied to the Monitoring of Tower Vibrations

Having shown the better performance of the stochastic approach with higher frequency sampling data, in this section we briefly discuss how the approach can be applied to specific monitoring of tower vibrations. The discussion is based on the approximation derived in Equation (8) and follows from the recent findings reported in Ref. [20].

The *Ansatz* of the stochastic modelling proposed above for tower vibrations, describes the evolution of short-time vibrations as the superposition of two contributions, being one deterministic (governed by function $D^{(1)}$), and the other stochastic (governed by function $D^{(2)}$). In parallel, tower vibrations result from the interplay of a mass of wind acting on the tower and the tower itself reacting back to the wind. Though there is no clear-cut match between the two contributions in our Ansatz and the physical reality of tower vibrations, we may regard the deterministic contribution ($D^{(1)}$) as the one describing the tower response to wind loads and the stochastic part ($D^{(2)}$) as the one describing the (turbulent) dynamics of wind. Consequently, from such an interpretation of the stochastic model introduced in Equation (3) we conclude that for monitoring normal functioning of tower vibrations we should focus on $D^{(1)}$.

As derived above in Equation (8), function $D^{(1)}$ can be interpreted as an elastic force, parameterized by a spring constant k. This spring constant characterizes the vibrating tower and therefore, in case structural defects occur, one expects that the vibrations will be qualitatively different from there on, showing a *different* $D^{(1)}$ in a way that a plot of it similar to Figure 5b will show a different slope, i.e., different value of the spring constant.

Before concluding, one additional point needs to be discussed. The absolute quality of the wind speed measurement is not crucial to our method. Most important for our method is the wind speed that the rotor is exposed to. In this sense it is even preferable to let the anemometer share the same tower oscillations as the rotor. However, we believe that the influence of tower oscillations on the wind speed measurement is small anyway. A very coarse estimation of nacelle oscillations leads to amplitudes in the order of 1 m at a frequency clearly below 1 Hz. This may give rise to wind speed measurement deviations below 1 m/s. As for our analysis we use wind speed classes of 0.5 m/s width, most of these tower oscillations would not have any effect. Moreover, these deviations will average out due to their periodic nature. As our analysis is of statistical nature, a significant effect is not expected.

5. Conclusions

We have presented a stochastic approach which is capable of modelling the normal behaviour of wind turbine tower acceleration for its monitoring. Our results demonstrate that, compared to neural network (NN) models and using a single input, the Langevin model is able to better reconstruct non-Gaussian fluctuations of signals. In general both models provide a good estimate for the central part of the original signal, but the stochastic approach also reconstructs its complete variance.

We have also shown that a normal behaviour model based on the stochastic approach provides more information about the original signal than one based on NNs. Still, one should emphasize that previous publications on normal behaviour models, to monitor vibration of wind turbines based on SCADA [12,13] using 10-s sampling signals, demonstrated the suitability of NNs and control charts as a method to monitor signals in time-domain. Our results for NN match the applications found in the literature, where they have been used with 10-min sampling time-series [10]. When using low-frequency data, the NN is as good as our stochastic approach.

Putting our results in perspective, we state that NNs are a good choice to predict average values or less fluctuating signals. For situations where 10-min data is not sufficiently, namely when dealing with cumulative loads for which the large wind fluctuations of short intervals play an important role, NNs seem to retrieve not so good results as the stochastic approach.

All in all, in those cases where the sampling rate is low, the NNs remain as a good option to take into account, as the monitoring is sufficient in time-domain and the requirements to apply the stochastic approach are not always fulfilled. In cases where the sampling rate of the signal is high

(1 Hz in our case), the stochastic approach is clearly a better choice. It allows for monitoring normal behaviour in time-domain as with NNs, but also extends the possibilities to the frequency-domain. As discussed above, to notice that a central advantage of the stochastic approach is that deterministic and noisy contributions can be separated leading to improvements in the statistical reproduction of the output fluctuations. The study of the long time behaviour of the deterministic part of the Langevin equation used in the stochastic approach can also provide information about the health of the structure, in a similar direction as previous work [20]. From this paper, one can now use the available routines [28,29] and extend the stochastic approach here applied to other wind turbine properties, in particular to other loads.

Acknowledgments: This work was partly funded by the German Federal Ministry of Economic Affairs and Energy and the State of Lower Saxony as part of the research project "Probabilistic load description, monitoring, and reduction for the next generation of offshore wind turbines (OWEA Loads)", grant number 0325577B, and also by the Ministry of Science and Culture of Lower Saxony in the project "Ventus Efficiens" (ZN3024). Additionally, financial support from the Deutsche Forschungsgemeinschaft (MA 1636/9-1) and the Open Access Publishing Fund of Osnabrück University is gratefully acknowledged. The authors also thank Senvion SE for providing the data here analyzed.

Author Contributions: Pedro G. Lind and Luis Vera-Tudela performed the simulations. Pedro G. Lind and Luis Vera-Tudela prepared the manuscript. Matthias Wächter, Martin Kühn and Joachim Peinke proposed the ideas and methodologies. All authors revised the text and output results.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Yang, W.; Tavner, P.J.; Crabtree, C.J.; Feng, Y.; Qiu, Y. Wind turbine condition monitoring: Technical and commercial challenges. *Wind Energy* **2014**, *17*, 673–693.
2. Kusiak, A.; Zhang, Z.; Verma, A. Prediction, operations, and condition monitoring in wind energy. *Energy* **2013**, *60*, 1–12.
3. Wang, K.S.; Sharma, V.S.; Zhang, Z.Y. SCADA data based condition monitoring of wind turbines. *Adv. Manuf.* **2014**, *2*, 61–69.
4. Tchakoua, P.; Wamkeue, R.; Ouhrouche, M.; Slaoui-Hasnaoui, F.; Tameghe, T.; Ekemb, G. Wind Turbine Condition Monitoring: State-of-the-Art Review, New Trends, and Future Challenges. *Energies* **2014**, *7*, 2595–2630.
5. Zaher, A.; McArthur, S.; Infield, D. Online wind turbine fault detection through automated SCADA data analysis. *Wind Energy* **2009**, *12*, 574–593.
6. Rahman, M.; Ong, Z.; Chong, T.; Julai, S.; Khoo, S. Performance enhancement of wind turbine systems with vibration control: A review. *Renew. Sustain. Energy Rev.* **2015**, *51*, 43–54.
7. Bogoevska, S.; Spiridonakos, M.; Chatzi, E.; Dumova-Jovanoska, E.; Höffer, R. A Data-Driven Diagnostic Framework for Wind Turbine Structures: A Holistic Approach. *Sensors* **2017**, *17*, 720.
8. Wang, J.; Heng, J.; Xiao, L.; Wang, C. Research and application of a combined model based on multi-objective optimization for multi-step ahead wind speed forecasting. *Energy* **2017**, *125*, 591–613.
9. Cruz Garcia, M.; Sanz-Bobi, M.A.; del Pico, J. SIMAP: Intelligent System for Predictive Maintenance. Application to the health condition monitoring of a windturbine gearbox. *Comput. Ind.* **2006**, *57*, 552–568.
10. Schlechtingen, M.; Ferreira Santos, I. Comparative analysis of neural network and regression based condition monitoring approaches for wind turbine fault detection. *Mech. Syst. Signal Process.* **2011**, *25*, 1849–1875.
11. Kusiak, A.; Verma, A. Analyzing bearing faults in wind turbines: A data-mining approach. *Renew. Energy* **2012**, *48*, 110–116.
12. Kusiak, A.; Zhang, Z. Analysis of Wind Turbine Vibrations Based on SCADA Data. *J. Sol. Energy Eng.* **2010**, *132*, 031008.
13. Zhang, Z.; Kusiak, A. Monitoring Wind Turbine Vibration Based on SCADA Data. *J. Sol. Energy Eng.* **2012**, *134*, 021004.
14. Bangalore, P.; Letzgus, S.; Karlsson, D.; Patriksson, M. An artificial neural network-based condition monitoring method for wind turbines, with application to the monitoring of the gearbox. *Wind Energy* **2017**, *20*, 1421–1438.

15. Lind, P.G.; Herráez, I.; Wächter, M.; Peinke, J. Fatigue Load Estimation through a Simple Stochastic Model. *Energies* **2014**, *7*, 8279–8293.
16. Russo, A.; Raischel, F.; Lind, P.G. Air quality prediction using optimal neural networks with stochastic variables. *Atmos. Environ.* **2013**, *79*, 822–830.
17. Müller, K.; Cheng, P.W. Validation of uncertainty in IEC damage calculations based on measurements from alpha ventus. *Energy Procedia* **2016**, *94*, 133–145.
18. Cheng, W.; Lutz, T.; Kühn, M.; Bartsch, J.; Kruse, J.; Schaumann, P.; Bange, J. *Probabilistische Lastbeschreibung, Monitoring und Reduktion der Lasten zukünftiger Offshore-Windenergieanlagen (OWEA Loads)*; Technical Report; WindForS and ForWind and Adwen GmbH and Senvion: Oldenburg/Stuttgart, Germany, 2017.
19. Mücke, T.; Wächter, M.; Milan, P.; Peinke, J. Langevin power curve analysis for numerical wind energy converter models with new insights on high frequency power performance. *Wind Energy* **2015**, *18*, 1953–1971.
20. Rinn, P.; Heißelmann, H.; Wächter, M.; Peinke, J. Stochastic method for in-situ damage analysis. *Eur. Phys. J. B* **2013**, *86*, 3.
21. Shmueli, G. To Explain or to Predict? *Stat. Sci.* **2010**, *25*, 289–310.
22. Borovicka, T.; Jirina, M.J.; Kordik, P.; Jirina, M. Selecting representative data sets. In *Advances in Data Mining Knowledge Discovery and Applications*; Karahoca, A., Ed.; InTech: Rijeka, Croatia, 2012; pp. 43–70.
23. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed.; Springer: New York, NY, USA, 2013; p. 759.
24. May, R.; Dandy, G.; Maier, H. Review of input variable selection methods for artificial neural networks. In *Artificial Neural Networks—Methodological Advances and Biomedical Applications*; Suzuki, K., Ed.; InTech Europe: Rijeka, Croatia, 2011; pp. 19–44.
25. Friedrich, R.; Peinke, J. Description of a Turbulent Cascade by a Fokker-Planck Equation. *Phys. Rev. Lett.* **1997**, *78*, 863.
26. Siegert, S.; Friedrich, R.; Peinke, J. Analysis of Data of Stochastic Systems. *Phys. Lett. A* **1998**, *243*, 275–280.
27. Friedrich, R.; Peinke, J.; Sahimi, M.; Tabar, M. Approaching complexity by stochastic methods: From biological systems to turbulence. *Phys. Rep.* **2011**, *506*, 87–167.
28. Rinn, P.; Lind, P.; Wächter, M.; Peinke, J. The Langevin Approach: An R Package for Modeling Stochastic Processes. *J. Open Res. Softw.* **2016**, *4*, e34.
29. Rinn, P.; Lind, P.G.; Bastine, D. Langevin 1D and 2D. 2015. Available online: https://cran.r-project.org/web/packages/Langevin/ (accessed on 28 October 2017).
30. Lehle, B. Analysis of stochastic time series in the presence of strong measurement noise. *Phys. Rev. E* **2011**, *83*, 021113.
31. Anvari, M.; Tabar, R.R.; Peinke, J.; Lehnertz, K. Disentangling the stochastic behaviour of complex time series. *Nat. Sci. Rep.* **2016**, *6*, 35435.
32. Zhang, Z.; Verma, A.; Member, S.; Kusiak, A. Fault Analysis and Condition Monitoring of the Wind Turbine Gearbox. *IEEE Trans. Energy Convers.* **2012**, *27*, 526–535.
33. Santos, P.; Maudes, J.; Bustillo, A. Identifying maximum imbalance in datasets for fault diagnosis of gearboxes. *J. Intell. Manuf.* **2015**, doi:10.1007/s10845-015-1110-0.