21th International Conference on Renewable Energies and Power Quality (ICREPQ'23)
Madrid (Spain), 24th to 26th May 2023
Renewable Energy and Power Quality Journal (RE&PQJ)
ISSN 2172-038 X, Volume No.21, July 2023

A method to assess the performance of small wind turbines

W.-G. Früh[1]

[1] Institute of Mechanical, Process and Energy Engineering
Heriot-Watt University, Riccarton, Edinburgh EH14 4A (Scotland, UK)
Phone/Fax number: +0044 131 451 4374, e-mail: w.g.fruh@hw.ac.uk

Abstract. While large wind farms are well monitored, with a wealth of data provided through a SCADA system, the only information about the behaviour of small wind turbines is often only through the metered electricity production. Given the variable electricity output, it is difficult to ascertain whether a particular electricity production in a metering period is the result of the turbine operating normally, or if a fault is resulting in a production less than possible. This paper presents a method to correlate metered electricity output from a set of 5 kW wind turbines with weather information from a weather station some distance from the turbine. That correlation will then be classified into 'expected' and 'unusual' performance using Principal Component Analysis.

Key words. Small Wind Turbines, Performance Assessment, Principal Component Analysis

1. Introduction

While industrial-scale wind energy projects using MW-scale turbines have grown rapidly over the last two decades, smaller wind turbines have grown at a much slower rate and somewhat stood in the shadow of the large wind project. However, they play an important role in supplying energy to smaller and rural communities, as well as for embedded generation in sub-urban and even urban environments [1], [2]. In terms of their cost-effectiveness, small wind turbines are at an inherent disadvantage over the large wind energy market, since they cannot exploit economies of scale while also operating closer to the ground where the wind is always lower than higher in the atmosphere. Therefore, it is even more important to design or select the most suitable turbines, the best location for them, and to ensure that they operate as well as they can.

This translates into three major specific challenges for small-scale wind: the first is related to the turbine design as the smaller wind turbines need to be designed for more turbulent and volatile wind conditions than larger turbines whose rotors are placed much higher in the atmosphere. While design of large turbines has largely converged to an industry standard of a pitch-controlled, variable-speed 3-bladed horizontal-axis wind turbine, the design of small wind turbines is still much more diverse [3]

The second challenge is the optimum micro-siting of turbines, as the local wind conditions vary significantly across a small area. For example, the performance of a roof-top mounted turbine varies substantially for different points on the roof [4]–[6].

The third challenge is in the monitoring and maintenance of smaller wind turbines. While large turbines are fitted with a large number of sensors to monitor environmental and turbine conditions [7], smaller turbines are often not monitored except through the electricity meter monitoring the electricity production accumulated over an accounting period, typically at half-hourly intervals or even less frequently.

The aim of this paper is to present a method on evaluating the performance of a set of small turbines using only the metered electricity production and wind speed and direction data from a 'near-by' weather station, where nearby could be several 10s of kilometres from the turbines. The method will be illustrated for an installation in Scotland.

2. Analysis methodology

The data available are metered half-hourly electricity production from an installation complemented by hourly

wind speed and direction data for the same time period from a nearby anemometer on a 10 m tall met. mast operated by the UK's meteorological office.

As there is no 'training' period available to judge good performance against below-expected performance, the method is here developed in two separate steps. In the first stage, an empirical performance curve of the wind turbine installation will be constructed through extracting performance distributions of the turbines in pre-defined wind speed bins of the corresponding wind speed data.

In the second step, Principal Component Analysis (PCA) will be applied to the data without any other explicit fitting. In this PCA analysis, a set of singular basis vectors is constructed through a PCA analysis of the subset of data which were identified as 'good turbine performance' according to the empirical performance curve. The entire set of data is then projected onto that singular basis vector space.

The argument here is that other 'good' behaviour will be mapped onto the same locations in that PCA space as the training data while data associated with unusual behaviour will be projected to other parts of that space. This means that the space spanned by the singular basis vectors can be sub-divided into regions with different levels of performance quality. The quality score of the region in which a particular measurement is mapped on to, therefore results in a quality measure of that particular turbine productivity information.

3. Sample Data

The wind energy data used to develop and illustrate the method were provided by the owner of an installation on an island in the north of the UK, while the wind resource information was provided from the nearest UK Met. Office weather station through the MIDAS dataset [8]. This data set provides hourly average wind speed and gusts in a 10-minute measurement period leading up to the measurement time stamp, both rounded to the nearest *knot*, and the wind direction rounded to ± 10° from an anemometer at a nominal height of 10 m above ground.

The installation consists of three 5 kW turbines with a nominal rated wind speed of 13 m/s. They are mounted on 15 m tall poles and have upwind horizontal-axis rotors, passively yawed through a vane. The site is located 25 km north of the weather station associated with the nearest airport.

Just over two years of data are available, from July 2013 to July of 2015, with one extended gap of 43 days in November and December 2013, and a shorter gap of 6 days in February 2014. The wind speeds recorded during the gaps do not show any unusual distribution, and certainly do not include any extreme events. As a result, the data gaps are inferred to be due to communications issues and not linked to the wind resource. With a data availability of 93.5 % and 95 % for the two data sets, it can be assumed that the gaps will not introduce a significant bias in the analysis, and are simply ignored in the analysis. This is possible, since the analysis is purely based on temporally independent samples of simultaneous measurements

The available installation performance data were only the metered electricity fed into the grid with a half-hour resolution. To align the data structure between the turbine performance and the wind resource, the electricity production from each half hour before and after a full hour were added to provide hourly production centred around the wind speed measurements. This resulted in a total set of 17 088 data triples of electricity production, wind speed and wind direction.

4. Empirical performance curve

Figure 1 shows a simple scatter plot of the electricity production against the raw wind speed data from the met. mast (in units *knot*). Figure 1 shows a clear correlation between wind speed and energy production, including clear evidence of cases where the output appears to be due to only one or two turbines operating instead of all three. In addition, there is substantial scatter around that easily understood variation.

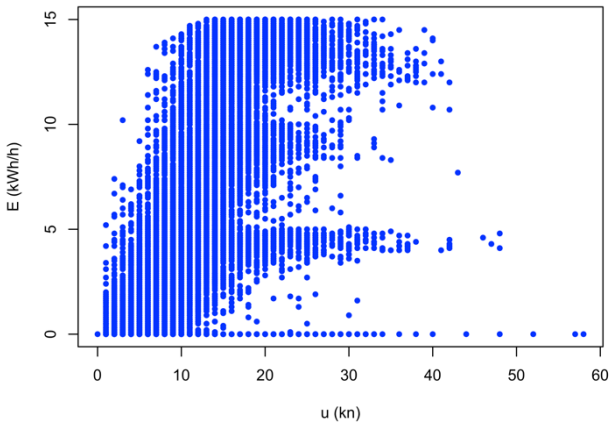

Figure 1. Metered electricity production against wind speed from the nearby weather station.

Given that the given wind speed is rounded to the nearest integer, the natural binning for the analysis is by integer *knots*. First, a box-and-whisker plot for the electricity production in each wind speed bin is created and shown in Figure 2, with a limit of the upper whisker set so that the upper quartile is no larger than half of the third quartile. This is to exclude outliers in the lower wind speed range. Given the different likelihood of a particular wind speed (consistent with a Weibull distribution), the number of samples in each wind speed bin varies substantially, from a few wind speed bins with single observations at the very high wind speeds, up to over 1 000 data points at the most likely range between 7 and 10 kn, with over 100 data points in each bin in the range between 2 and 29 kn.

From this, the equivalent of the 'rated wind speed' is determined as the first bin where the upper whisker exceeds 95% of the installed capacity. This is used to determine a scaling between the measured wind and the

reference wind speeds from the manufacturer's performance curve, and the identified empirical rated wind speed is highlighted by the red dot and lines in Figure 2. With three turbines, the threshold for selecting the rated wind speed is 14.25 kWh. The boxplot at 14 kn is the first where that hourly electricity production is no longer regarded as an outlier.

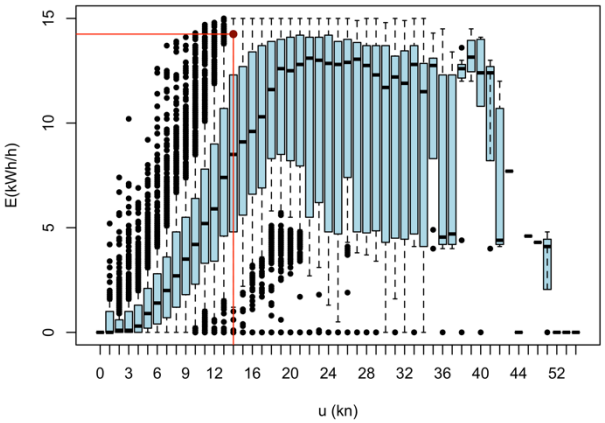

Figure 2. Box-and-whisker plot of the observed electricity production against weather data in wind speed bins of size 1 kn.

In the absence of confirmed 'good' behaviour of the installation, the empirical performance curve will be defined as the best 3/8th of the data. The resulting range of performance one can expect from the installation is shown in Figure 3, as the area shaded in blue denoting the range, and the solid line for the mean of the data within that range. Of the total 17 088 measurements, 8 698 are within the blue band. Overlaid on this is the manufacturer's performance curve mapped onto the best estimate of the wind speed at the site, as the dashed red line. It is clear that the installations appear to perform better at low wind speed but that the performance increases much more gradually than a single turbine under controlled conditions. Furthermore, even though the turbine manufacturer states that the rated power should be achievable up to a wind speed of 60 m/s (or 117 kn), the installed turbines do show a gradual but clear drop of productivity at high wind speeds. The reason for this discrepancy is not clear.

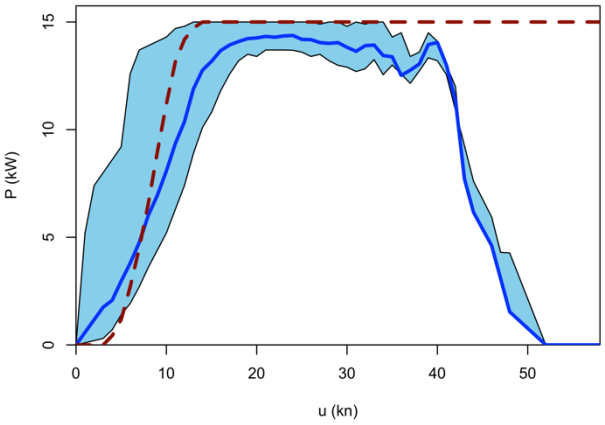

Figure 3. Empirical performance curve for the installation against the regional weather data. The dashed red line indicates the manufacturer's performance data for a single turbine.

This can be further refined into a directional sensitivity by plotting the median performance by wind speed and wind direction bins, shown as a heat map in Figure 4. There it is clear that a systematic directional sensitivity is most clearly observed just below the empirical rated wind speed. For wind directions around 70° to 90° and 220° to 270°, the installation reaches capacity at low wind speeds than for the other wind directions. Without reference to local maps, it is impossible to say whether this is due to turbine wakes or topographic effects.

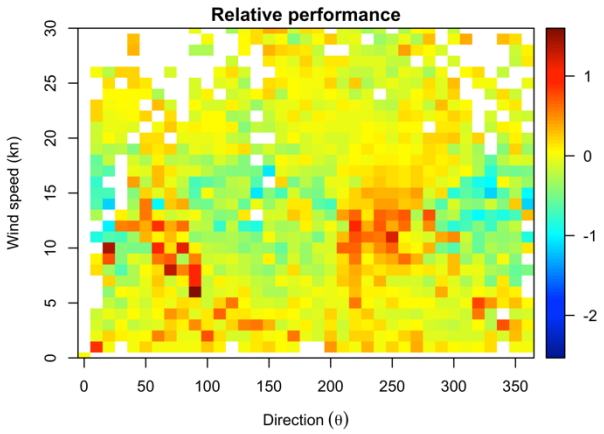

Figure 4. Heat map of variation of mean performance by wind direction. Red colours indicate enhanced productivity, blue colour indicate reduced productivity.

5. PCA-based performance estimation

For the development of the PCA-based performance estimation, the 8 698 observations within the blue range in Figure 3 are then taken to train the PCA estimator. Principal Component Analysis is a widely used technique for data reduction, e.g. [9], [10], phase space reconstruction [11], forecasting [12] and wind resource assessment [13]. It is usually recommended to scale the input variable so that no variable introduces a bias due to a much larger range. The procedure starts with creating a data matrix $\mathbf{Y}$, consisting of column vectors of simultaneous data, $\mathbf{Y}^{i,j=1,2,3} = (E_i, u_i, \theta_i)$, with the electricity production E_i at time point i, the wind speed u_i and the wind direction θ_i. All data were normalised to be within the interval of 0 to 1. This means that full electricity production of 15 kWh is rescaled to 1, wind speeds are divided by the highest observed wind speed of 58 kn, and wind direction is now from 0 northerly winds through 0.25 for easterly winds to 1 for 350°.

The PCA itself is based on the Singular Value Decomposition (SVD) of that matrix $\mathbf{Y}$, such that

$$\mathbf{Y} = \mathbf{P}\,\mathbf{\Lambda}\,\mathbf{\Sigma} \tag{1}$$

with

- $\mathbf{P}$ the matrix with column vectors of 'principal components', which give the location in the PCA space. The matrix is orthonormal: each column

vector has unit variance and each column vector is orthogonal to all other column vectors

- **Λ** the diagonal matrix of 'singular values' which measure how much the variation in the direction of that basis vector contributes to the overall variance,

- **Σ** the matrix with column vectors defining the basis vectors, often referred to as 'singular vectors'. This matrix is also orthonormal. (Note that some software package implement the SVD implemented algorithm to return the transpose, i.e. the basis vectors are the rows).

The new representation of the data in the PCA space shows the 'coordinates' of a measurement as the Principal Component value against the axis of the associated basis vector. The data from the training set in this space are illustrated at a projection of the points onto the plane spanned by the first two basis vectors, i.e., PC1 versus PC2, in Figure 5a while Figure 5b shows the data in the plane spanned by the second and third direction, i.e., PC3 vs PC2. Variation in PC1 is largely due to the nominal performance characteristics while PC2 and PC3 capture the variability around that. Therefore, the plane PC2-PC3 is that to judge performance against.

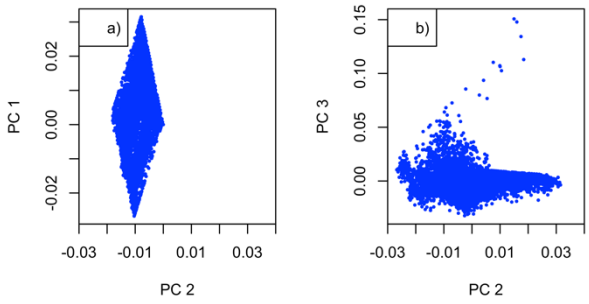

Figure 5. Scatter plots of a) the first Principal Component values against the second, and b) the third against the second for the training data.

The scatter plot in Figure 5b is then used to create a density map for individual area elements in that space. Elements without any data points are assigned a density of zero, the element with the maximum number of points a density of 1, and all other areas are assigned the ratio of the local number of points by the maximum. This results in the heat map shown in Figure 6. Assuming (for the purpose of this demonstration) that the most reliable data are the most frequent, this density map can now be used as a quality index to estimate the quality of new measurements against these reference data.

Mapping all (or any future measurements) onto this reference quality map follows the same principle as equation (1) but instead of finding **P**, **Λ**, and **Σ**, now the task is to calculate new **P** values with the singular values and vectors from the training set and the new data, $\mathbf{Y}'$:

$$\mathbf{P}' = \mathbf{Y}'\boldsymbol{\Sigma}^{T}\boldsymbol{\Lambda}^{-1} \tag{2}$$

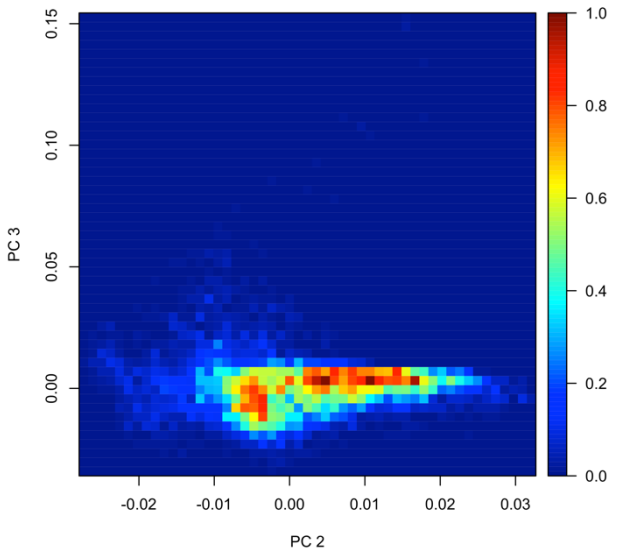

Figure 6. Density map of the installation's 'good' behaviour in Principal Component space.

where the inversion of the singular vectors is given by the transpose of the matrix (since it is orthonormal), and the inversion of the singular value matrix is the inverse of each diagonal element (since it is diagonal). Applying this mapping to the full data set results in Figure 7 where the colour of the points is based on the local quality index, with light blue having an index of 1, and red has an index of 0.

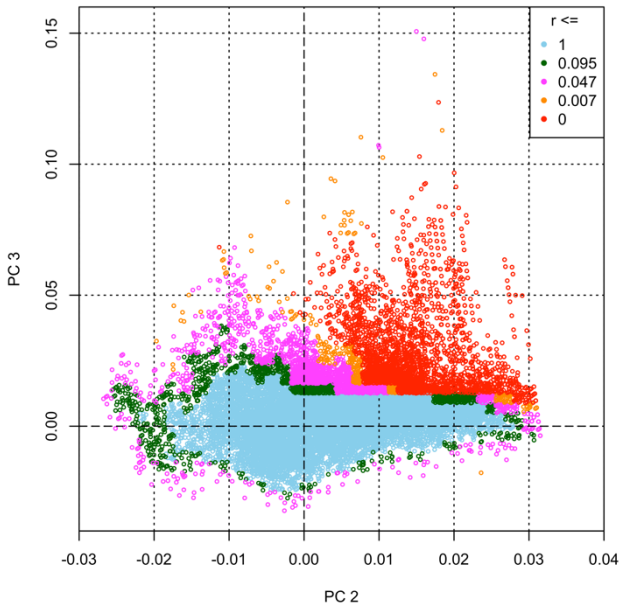

Figure 7. Mapping of all data on to the Principal Component space, with colour coding by quality score, *r*.

Converting the colour-coded principal components back into the more familiar electricity-wind speed graph in Figure 8 confirms that the method has successfully identified the main branch of the empirical performance curve, and scored the points lower, the further away they are from the reference curve.

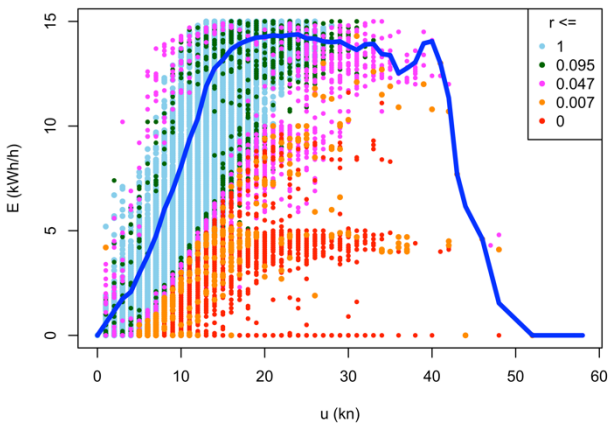

Figure 8. Installation performance, with performance colour-coded by quality score, *r*.

6. Discussion

Figure 8 demonstrates that the performance monitoring method is clearly able to map new data onto the reference information created as a benchmark of 'good performance'. Furthermore, the computational demands are very small: on a 1.6 GHz Dual-Core Intel i5 laptop, the creation of the empirical performance curve, shown in section 4, was completed within 5 seconds including reading in the data files using just under 1 second of CPU time. The PCA of the reference data was completed in 2.7 s using 1.7 s of CPU, and the mapping of the entire data set onto the reference singular vectors took 1.4 s and 0.6 s of CPU time. As such, it shows clear potential to be an easily implemented tool to provide monitoring of small wind turbines.

However, Figure 8 also shows that many of the intuitively accepted points at higher wind speeds and near the installed power have a lower quality index than the bulk of the data at lower wind speeds. This can be traced back to the original approach of creating the reference PCA using the actual filtered observations. Since the most common wind speeds are in the range of 3 to 20 kn, the density map in Figure 6 will necessarily have the highest value, therefore introducing a bias towards most common data as opposed to data trusted by performance. While this does not invalidate the PCA approach in itself, it points to a source of bias in the construction and selection of the reference data to train the PCA with.

One suggestion to explore in future work is to create surrogate reference performance data based on the reference curve constructed in Figure 3 but, instead of using the available data within the acceptable range delineated by the shaded area, the surrogate reference data would take the observed distribution within each wind speed bin but then populate that bin with randomly generated performance data which conforms with the observed distribution but ensures that each wind speed bin has the same number of samples. This would preserve the shape of the reference performance curve but remove the bias from the fact that some wind speeds are more common than others.

Irrespective of this current weakness, it is instructive to apply the method to judge the performance of the installation against expectations. Figure 9 creates the data for this, with the performance shown as the capacity factor. The actual performance is the central red column, while the first column on the left is the expected performance when using the manufacturer's performance curve (the dashed red line in Figure 3) and the best estimate of the local wind speed. It is clear that the manufacturer's information would lead to a significant over-estimation of the yield, followed by a disappointment of the client. Testing how realistic (or not) that manufacturer's estimation is, can be done by using the performance curve shown as the solid blue line in Figure 3. Given that it outperforms the manufacturer's curve at the very common lower wind speeds, the potential is actually slightly higher even than promised by the manufacturer. This suggests that the actual performance falls not only well below expectations but also below its potential.

To estimate how much of that potential is realistically recoverable by regular monitoring, the data below expectations have been put into two categories. The first category is likely to be related to the grid infrastructure and therefore beyond the control of the owner. Since it is unlikely that all three turbines fail at the same time, zero output despite sufficient wind is assumed to be caused by grid constraints and not turbine faults.

The second category with the remaining data below expectations could be improved by regular monitoring and maintenance. By applying the empirical performance curve to those data, the additional electricity production can be estimated, shown in the two columns on the right. They show a clear potential for significant improvement, while also showing that the potential purely given by the wind resource will not be achievable. As a result, this method not only allows to identify sub-optimal performance but also provide a way to estimate a more realistic productivity than just using the mean wind speed and manufacturer's data sheets.

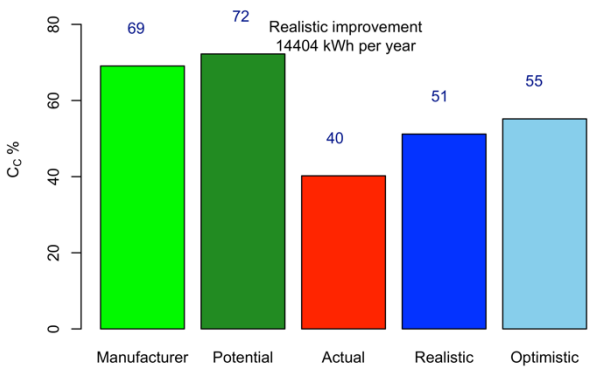

Figure 9. Comparison of actual performance to potential performance.

7. Conclusions

This paper has developed a method to evaluate the performance of small renewable installations which are often minimally monitored, only through logging the electricity production without any resource information or device health monitoring. In the absence of a gold standard to train the performance evaluation against, an initial analysis was completed to provide a benchmark. This was then used to not only identify performance data falling outside that benchmark but also to assign a quality score.

As such, this method appears to be robust and applicable to a range of small installation, including solar PV systems. However, further development and validation of the methods is required, partly against more wind generators partly against other generation technologies, but also a more robust method in the creation of the benchmark is required. This could either be done by in-situ measurements – but that would come at an operational cost which may be beyond the operator's means. Alternatively, a more robust method to infer the benchmark is proposed which would address the current bias introduced by the different likelihood of certain weather conditions prevailing. This method is identified as the key next step to develop this method further.

Acknowledgement

We would like to thank the UK Meteorological office for providing access to the wind data from the MIDAS record through the CEDA Data Archive (https://archive.ceda.ac.uk/) and for providing the additional anemometer details.

References

[1] Z. Simic, J. G. Havelka, and M. Bozicevic Vrhovcak, "Small wind turbines - A unique segment of the wind power market," *Renew. Energy*, vol. 50, pp. 1027–1036, 2013, doi: 10.1016/j.renene.2012.08.038.

[2] A. KC, J. Whale, and T. Urmee, "Urban wind conditions and small wind turbines in the built environment: A review," *Renew. Energy*, vol. 131, pp. 268–283, 2019, doi: 10.1016/j.renene.2018.07.050.

[3] A. Bianchini *et al.*, "Current status and grand challenges for small wind turbine technology," *Wind Energy Sci.*, vol. 7, no. 5, pp. 2003–2037, 2022, doi: 10.5194/wes-7-2003-2022.

[4] Energy Saving Trust, "Location, location, location: Domestic small-scale wind field trial report," 2007.

[5] M.A. Heath, J.D. Walshe, and S.J. Watson, "Estimating the potential yield of small building-mounted wind turbines," *Wind Energy*, vol. 10, no. 3, pp. 271–287, 2007, doi: doi.org/10.1002/we.222.

[6] L. Ledo, P.B. Kosasih, and P. Cooper, "Roof mounting site analysis for micro-wind turbines," *Renew. Energy*, vol. 36, no. 5, pp. 1379–1391, 2011, doi: 10.1016/j.renene.2010.10.030.

[7] K. Kong, K. Dyer, C. Payne, I. Hamerton, and P. M. Weaver, "Progress and Trends in Damage Detection Methods, Maintenance, and Data-driven Monitoring of Wind Turbine Blades – A Review," *Renew. Energy Focus*, vol. 44, 2022, doi: 10.1016/j.ref.2022.08.005.

[8] UK Meteorological Office, "Met Office Integrated Data Archive System (MIDAS) Land and Marine Surface Stations Data (1853-current). NCAS British Atmospheric Data Centre." 2012.

[9] B. Everitt and G. Dunn, *Applied Multivariate Data Analysis*, 2nd ed. Arnold, 2001. doi: 10.1002/9781118887486.

[10] A. Field, J. Miles, and Z. Field, *Discovering Statistics using R*. London, Thousand Oaks, New Delhi, Singapore: SAGE Publications Ltd., 2013.

[11] D. S. Broomhead, R. Jones, G. P. King, and E. R. Pike, "Singular systems analysis with application to dynamical systems," in *Chaos, noise and fractals*, E. R. Pike and L. A. Lugiato, Eds. Bristol: Adam Hilger, 1987, pp. 15–27. doi: 10.1201/9781003069553-2.

[12] C. Skittides and W.-G. Früh, "Wind forecasting using Principal Component Analysis," *Renew. Energy*, vol. 69, pp. 365–374, 2014, doi: http://dx.doi.org/10.1016/j.renene.2014.03.068.

[13] C. Skittides and W. G. Früh, "A new Measure-Correlate-Predict Wind Resource Prediction method," in *International Conference on Renewable Energies and Power Quality (ICREPQ'15)*, 2015, vol. 13, no. 13. doi: 10.24084/repqj13.4255.