Review

A Review of Approaches for the Detection and Treatment of Outliers in Processing Wind Turbine and Wind Farm Measurements

Mingzhe Zou and Sasa Z. Djokic *

School of Engineering, the University of Edinburgh, Edinburgh EH9 3DW, UK; mingzhe.zou@ed.ac.uk

* Correspondence: sasa.djokic@ed.ac.uk

Received: 14 July 2020; Accepted: 12 August 2020; Published: 15 August 2020

Abstract: Due to the significant increase of the number of wind-based electricity generation systems, it is important to have accurate information on their operational characteristics, which are typically obtained by processing large amounts of measurements from the individual wind turbines (WTs) and from the whole wind farms (WFs). For further processing of these measurements, it is important to identify and remove bad quality or abnormal data, as otherwise obtained WT and WF models may be biased, or even inaccurate. There are wide ranges of both causes and manifestations of these bad/abnormal data, which are often denoted with the common general term “outlier”. This paper reviews approaches for the detection and treatment of outliers in processing WT and WF measurements, starting from the discussion of the commonly measured parameters, variables and resolutions, as well as the corresponding requirements and recommendations in related standards. Afterwards, characteristics and causes of outliers reported in existing literature are discussed and illustrated, as well as the requirements for the data rejection in related standard. Next, outlier identification methods are reviewed, followed by a review of approaches for testing the success of outlier removal procedures, with a discussion of their potential negative effects and impact on the WT and WF models. Finally, the paper indicates some issues and concerns that could be of interests for the further research on the detection and treatment of outliers in processing WT and WF measurements.

Keywords: measurement data; outlier; wind turbine; wind farm

1. Introduction

Over the last few decades, there has been a significant increase in the penetration of wind-based electricity generation systems, with installed powers ranging from a few kilowatt (kW) individual wind turbines (WTs), to gigawatt (GW)-scale wind farms (WFs). As the number of wind energy conversion systems is increasing, so is the importance of having accurate information on their power outputs, which is difficult to obtain due to the strong and inherently stochastic variations of the wind energy resource, as well as the complexity of WT control systems and the highly dynamic nature of WT operation. For example, the assessment of uncertainties in the outputs of wind-based generation systems [1–3] is of importance in a number of power system applications, including condition monitoring and fault diagnosis [4], unit commitment and economic dispatch [5], as well as the analysis of system security, reliability and voltage stability [6].

The common approach for modelling WTs and WF is to use power curves (PCs) [7–10], which typically provide information on the generated power output (P_{out}) as a function of the input wind speed (WS). These PC-based WT and WF models have numerous applications, e.g., in wind power forecasting [11,12], WT condition monitoring [7,13], WF condition monitoring [14,15], wind energy resource assessment [16–18] and general networks analysis [19]. Manufacturers’ power curves (Mfr-PCs)

are typically used if no field (measured) data are available, as is the case in a planning phase, or general design phase of a WF. However, since Mfr-PCs are obtained in air-tunnels, i.e., under controlled conditions [20,21], they cannot reflect many WT and WF application-specific factors, such as the impact of the terrain, wake effects and the dynamics of input wind energy variations [22–25]. Furthermore, after several years of operation, natural ageing will result in a deterioration of WT performance from the originally specified Mfr-PC, which may be further strongly impacted by the meteorological conditions at locations where the WT is installed [10,24,26,27]. Accordingly, the actual field performance of an operational WT can be more accurately assessed from the equivalent operational PC-based WT model, which is typically derived from the available measurement data. For that purpose, the corresponding IEC Standard [21] provides guidance for power performance measurements of an electricity-producing WT, including requirements for the test equipment, measurement procedures, derivation of results, data normalisation and reporting format. This standard is specifically intended for the analysis of individual WTs. Although it does not specify nor instruct how an equivalent PC model could, or should be obtained for the whole WF, it does specify the parameters impacting the wind power output. These parameters should be taken into consideration during the measurement and modelling, such as wind direction (WD), air density (mainly influenced by temperature, air pressure and relative humidity), turbulence intensity and some other conditions (e.g., precipitation, icing, dirt and insect accretions, etc.) [21,23]. The PC-based model obtained from synchronous W_S and P_{out} measurements of an individual WT is in this paper denoted as a WT-level model (see e.g., [7,28]), while a PC-based model of the whole WF is denoted as a WF-level model (see e.g., [20]).

As with all measurement-based models and data-driven modelling approaches, the quality of the data used for both WT-level and WF-level modelling has to be checked [29]. The bad or abnormal data must be identified and removed (or suitably processed), as otherwise the obtained WT and WF models may be biased, or even inaccurate. There are wide ranges of both causes and manifestations of these bad/abnormal data, which are often denoted with the general term “outlier”. In the context of the PC-based WT and WF models, outliers are often identified as the specific measurement data that are far away from a much larger and closely clustered set of normal or regular measurement data, against which outliers are evaluated.

Correct identification and processing of outliers in the WT and WF measurements are particularly important in the condition monitoring and fault diagnosis applications [13,25,27,30–35]. For example, the presence of outliers is one of the biggest difficulties in applying appropriate PC-based monitoring methods, as the outliers usually result in unsatisfactory power curve limits and it is important that input data for the limit-setting algorithms do not include outliers [31]. Undetected or incorrectly processed outliers also have strong negative impact in applications related to wind power forecasting, power flow studies and economic dispatch analysis, since the models with outliers are likely to have lower accuracy, higher forecasting errors and suboptimal unit commitment results [36–39].

It can be generally concluded that accurate measurement-based models require one to separate outliers (i.e., bad quality or abnormal data) from the regular data (i.e., high quality or normal data), in order to correctly analyse and represent the underlying relations among the all considered model parameters, variables and dimensions. The importance of outlier analysis is reflected in a recent increase of interest of both scientific and engineering communities in approaches and methods for detection and processing of outliers in measurement-based WT and WF models. Most of the proposed approaches are statistical [14,15,18,20,23–26,29,31,33–35,37,39–49]. More specifically, Refs. [14,35] have classification-based methods, Refs. [15,20,33,39,40,44–47] present clustering-based models, Refs. [41,42,48] use correlation-based approaches, Refs. [18,24,25,29,43,44] apply normal-based, quartile-based or mean absolute deviation-based criteria, Refs. [26,31,34] propose recursive or iteration-based algorithms, Ref. [23] distinguishes the outliers by statistically robust data fitting, Ref. [37] utilises data smoothing and averaging followed by data censoring rule, Ref. [29] applies change point detection, Ref. [49] leverages computer visual technique, and [44] employs hypothesis test methods. Some other methods are based on physical characteristics and constraints [7,27,50], or on

combining both statistical approaches and physical constraints [51]. Finally, some of the work tried to develop so-called robust statistical models that could reduce the impact of outliers without their detection and elimination [30,32,52]. In terms of outlier treatment, in most of the work outliers are simply removed, e.g., [7,14,15,18,20,23,24,26,27,29,31,34,35,37,40,41,43,45,50,51], while few references treated outliers as missing values and tried to recover them, e.g., [42,48,53]. However, and despite the significant work on the detection, treatment and analysis of outliers in WT and WF measurement applications, there is still lack of the commonly accepted, or recommended methodologies for their comprehensive and systematic analysis, including clear definitions, comprehensive specification of their characteristics and confident identification of their causes and origins. This paper provides a review of the approaches and methods reported in available literature for the detection and treatment of outliers when processing WT and WF measurement data, as well as further discussion of some general and specific aspects of the related analysis of outliers.

After this introductory section, Section 2 provides a review of the most commonly used measurement data, parameters, variables and resolutions in WT and WF applications and also discusses corresponding requirements and recommendations from [21], which is an important standard for measurement-based performance evaluation of WTs. Section 3 reviews definitions, characteristics and causes of outliers and provides some illustrative examples, as well as the requirements for data rejection in [21]. In Sections 4 and 5, outlier identification methods are reviewed, followed by a review of approaches for testing the success of applied outlier treatment procedures and discussion of their potential negative effects and impact on the PC-based WT and WF models. Finally, Section 6 lists main conclusions and indicates some issues and concerns that could be of interests for the further research on the detection and treatment of outliers in processing WT and WF measurements.

2. Wind Turbine and Wind Farm Measurements

Most of the previous related work [7,13,18,23–29,31,34,35,39,41–43,45,47,49,51,53] focused on individual wind turbines (WT-level analysis), taking into consideration only the measurements from the sensors typically located at the WT nacelle. Some work, however, aimed to model the whole wind farm (WF-level analysis), when either: (a) data representing the whole WF are obtained directly, e.g., as a single-sensor measured total WF power outputs (P_{out}) data, typically obtained at the WF grid connection point, as well as a single-sensor measured WS and WD data, typically from the meteorological (MET) masts and towers [37,40,46,48], or (b) available measurements from individual WTs in a WF are in some way aggregated to derive the information on the whole WF indirectly [14,15]. An overview is provided below:

- Analysis of individual WT performance, using single WT measurements: [7,13,18,23–29,31,34,35,39,41–43,45,47,49,51,53];
- Analysis of whole WF performance, using single-sensor WF measurements: [37,40,46,48];
- Analysis of whole WF performance, using aggregation of measurements of all WTs in the same WF, i.e., not using directly measured single-sensor WF data [14,15]. For example, Ref. [14] applies principal component analysis (PCA, [54]) on the measurements from 89 individual WTs in a WF, which are processed to reduce dimensionality and obtain an equivalent WS for the whole WF, while equivalent P_{out} WF values are calculated from the P_{out} measurements of 89 WTs. In [15], average values of P_{out} , WS, rotor speed and blade pitch angle are collected from 22 individual WTs and then used to obtain three PC models for the whole WF: (i) power curve model (P_{out} -WS relationship), (ii) rotor curve model (rotor speed-WS relationship), and (iii) blade pitch curve model (blade pitch angle-WS relationship);
- Analysis of both WT and WF performance: [20,50]. For example, a data set consisting of 20 WTs in a WF is used in [20] to propose an outlier elimination approach for WT measurements, which is then applied to the WF's P_{out} -WS scatter-plot data, without specifying the method for obtaining the WF measurements (an example illustrating WF's P_{out} -WS scatter-plots can be found in Section 3.3). The measurements of both individual 274 WTs (P_{out} , WS and turbine availability

binary data, which are sampled every 10 min) and of the whole WF (WS, WD, P_{out} , etc.) are used in [50], with a focus on the whole WF performance. Individual WT availability data are used to help identification of outliers in the WF measurements, underlining the difference between the aggregated WF P_{out} (from the individual WTs' P_{out} data) and single-point measured whole WF P_{out} , which is attributed to losses.

In the WT and WF outlier detection and treatment research in available literature, WS and P_{out} are the two most often considered variables, as the impact of other meteorological variables is less significant. In cases where measurements of other variables were available, these extra data were either removed, or used only for some limited analysis, often without explicitly checking the presence of outliers in the extra data [13,14,18,20,25,26,28,29,31,35,37,40,41,43,45,46,48,49,51]. If other/extra variables are measured and considered in the outlier-related work, this was usually done at the data pre-processing stage, with examples discussed below:

- Measurements of rotor speed are added as an additional variable in order to help detection of outliers in WS and P_{out} measurements in [42], while WD is considered as an auxiliary variable in the model proposed for filtering outliers in P_{out} -WS data in [23].
- Measurements of additional variables are used for the transformation of P_{out} or WS data, e.g., based on guidance in [21] for data normalisation: in [24], temperature, humidity and air pressure are used to obtain the air density and then to transform P_{out} data; similar approach is applied in [50] to transform WS data. In [23,34], the same WS or P_{out} corrections as in [24,50] are used to eliminate the impact of the variance of WS fluctuations by expanding the PC into a Taylor series with the higher order terms neglected. Another example is use of turbulence intensity, as defined in [21], which has significant impact on ten-minute averaged WS and P_{out} measurements (see Section 3.2.2). Typically, the increased turbulence intensity leads to an increase of power output at the lower wind speeds ("ankle" of the power curve), while at the higher wind speeds ("knee" of the power curve), increased turbulence intensity results in a decrease of power output. Accordingly, a procedure for normalising PC data to a reference turbulence intensity from [21] is used in [55]. In that paper, a simulation-based approach and normal distribution (location parameter is the average wind speed within 10 min, scale parameter is the standard deviation of the wind speed within 10 min) of WS values within a 10-min window (and some other assumptions, for example, that at each instant the WT follows a zero turbulence PC) are used instead of the Taylor-series approach [23,34] (both approaches are presented in [55,56]).
- Examples of use of additional data where efforts are also made to detect potential outliers in these measurements are: [7], where the importance of ambient temperature and wind direction is emphasised, as temperature has the biggest influence on air density and, in turn, P_{out} (up to 20%) and [15,39,47], where various multivariate outlier detection approaches were applied on all considered variables (P_{out} , WS, rotor speed, blade pitch angle, etc.), as well as [27,50], which tried to detect outliers by analysing them separately in one dimension and simultaneously in multi-dimensional space. Finally, Ref. [53] does not explicitly present the outlier detection method, but it employs several parameters (P_{out} , WS, WD, rotor speed, gear temperature, blade pitch angle) and treats the outliers in these parameters as missing values.

In the above references, most of the measurements are sampled and collected at higher frequencies (usually 1 Hz), but are stored as 10-min average values [21], while only [13,18], and [25,26,31] used 5-min average values and [40,48] used 15-min average values.

All above discussed references use P_{out} and WS measurements and propose detection methods for potential outliers. However, five additional references are found with related, but inherently different analysis:

- The outliers are analysed in the measured high frequency vibration data (sampling frequency equal to 25.6 kHz) for health condition monitoring in [33], where they are collected via an accelerometer installed on the drivetrain structure of a WT.

- The outliers are considered for cyber-security analysis of false data injection attacks for two WFs (each with 18 WTs) using real-world hourly WS information in [44], but P_{out} values of WTs are calculated with a deterministic piecewise linear power curve.
- Robust statistical techniques that may reduce the effects of outliers and could be used as nominal baselines are proposed in [30,32,52], but the specific outlier detection methods are neither stated, nor used. Ref. [30] uses 12 kHz vibration data of WT bearings; Ref. [32] analyses 10-min average WS, P_{out} and blade bending moment data of WT; Ref. [52] builds models using 5-min average WF P_{out} data. These models are reviewed in more detail in Section 4.

The above discussion is summarised in Table 1 below.

Table 1. Summary of wind turbine and wind farm measurements in available outlier-related literature.

Ref.	Outlier-Related Research Focus	Measurement Data Considered in Outlier Analysis	Measurements Resolution
[29]	WT-level	WS, P_{out}	10-min
[41]	WT-level	WS, P_{out}	10-min
[42]	WT-level	WS, P_{out} , rotor speed	10-min
[7]	WT-level	WS, P_{out} , WD, ambient temperature	10-min
[35]	WT-level	WS, P_{out}	10-min
[43]	WT-level	WS, P_{out}	10-min
[53,57]	WT-level	WS, P_{out} , WD, rotor speed, gear temperature, blade pitch angle	Unspecified
[24]	WT-level	WS, P_{out} , ambient temperature, relative humidity, air pressure	10-min
[28]	WT-level	WS, P_{out}	10-min
[13]	WT-level	WS, P_{out}	5-min
[26]	WT-level	WS, P_{out}	5-min
[49]	WT-level	WS, P_{out}	10-min
[45]	WT-level	WS, P_{out}	10-min
[25]	WT-level	WS, P_{out}	5-min
[23]	WT-level	WS, P_{out} , WD, ambient temperature, relative humidity, air pressure	10-min
[34]	WT-level	WS, P_{out} , ambient temperature, relative humidity, air pressure	10-min
[18]	WT-level	WS, P_{out}	5-min
[31]	WT-level	WS, P_{out}	5-min
[51]	WT-level	WS, P_{out}	10-min
[47]	WT-level	WS, P_{out} , ambient temperature, rotor speed, generator speed, blade pitch angle, etc.	Unspecified
[39]	WT-level	WS, P_{out} , blade pitch angle	10-min
[27]	WT-level	WS, P_{out} , generator speed, nacelle temperature, etc.	10-min
[40]	WF-level	WS, P_{out}	15-min
[48]	WF-level	WS, P_{out}	15-min
[37]	WF-level	WS, P_{out}	Unspecified
[46]	WF-level	WS, P_{out}	Unspecified
[14]	WF-level	WS, P_{out}	10-min
[15]	WF-level	WS, P_{out} , rotor speed, blade pitch angle	10-min
[20]	WT and WF-level	WS, P_{out}	Unspecified
[50]	WT and WF-level	WS, P_{out} , ambient temperature, relative humidity, air pressure	10-min

3. Definitions, Characteristics and Causes of Outliers

In related literature, different terms and terminology are used to refer to, or to denote an outlier. The following three general cases are found: (1) the authors indicate the alternative term by using an abbreviation for interchangeability “i.e.,” between it and the term outlier; (2) the authors explicitly state that the specifically termed set of data are considered as the outliers; (3) the authors effectively treat the considered set of data as outliers by, for example, detecting them and either correcting or removing them from the further analysis. The alternative terms commonly found in the literature to denote an outlier are: “abnormal data” [29,43,45,46,49], “(data) error” [14,46], “incorrect data” [40], “unnatural data” [40], “irrational data” [40], and “wrong measurement” [24]. Furthermore, the same term may be used in distinctively different contexts in different references. For example, the term “missing data” is distinguished from the term outlier in [40,53], while the term “missing data” is used to refer to the data obtained when the communication system is faulty and when the last reading

before the fault has occurred was recorded over a longer period, leading to the series of constant or zero values in the recorded data in [47].

3.1. Definitions of Outliers

Most of the references define outliers in a statistical sense, as the data that are simply different or stand apart from the other data points, also deviate from the normal or logistic trends. In terms of PC-based models, outliers are points assumed to be outside the range of normal output characteristics of the power curve. In related literature approaching outliers from the statistical point of view, outliers are data points which usually appear to be under-power points, stopping points, or sparsely located points (e.g., either low wind speeds, but high power outputs, or high wind speeds, but low power outputs), which are usually identified in the P_{out} -WS scatter plot [29,43,44,46,48]. Instead of relying on P_{out} -WS scatter plot, [52] defines outliers by looking at time series, considering the recurrent and large changes in wind power time series as outliers, although it also states that these recordings can be frequently found.

There are a few references that define outliers in different ways, e.g., looking at operational WT performance. For example, [34] defines outliers as the values that are not corresponding to normal operating conditions, or when the WT is not available over the whole 10-min averaging period. Reference [34] does not provide an explicit definition of normal operating conditions, but according to [23] asserts that the WT will work properly during the 10-min period if the power data points are averaged considering the power production; otherwise, the recorded data do not represent the normal operation of the WT. Although not giving clear definition of outliers, [27,32] state that many outliers are particularly likely to be generated during the (extreme) transient operations, such as the start-ups and shut-downs of the WT. However, it is not uncommon for a WT to have several start-up and shut-down operations within a 10-min period, especially when the wind speed and/or wind turbulence is high, and such approaches will remove a number of rather usual operational points from the WT model.

The most general and intuitive approach in analysing outliers is that an outlier is denoted as the unusual or unexpected value, which is quite different from the set of values obtained or sampled under the same or very similar conditions (sometimes called inliers). If an outlier is related to a measured or in some other way observed value, its presence is typically assessed against a larger and clustered set of observations or measurement points obtained for the same or similar conditions. These points are used to both establish acceptable range(s) of variations (e.g., due to a limited measurement accuracy) and, through it, a statistically significant reference set(s) of values for the evaluation of outliers. Basically, this represents identification of outliers from a statistical point of view, where outliers are values that are at a large distance (or isolated) from the corresponding reference set of values (as denoted in e.g., [25,51]). In a different, but still associated context, results of measurements and observations are often parts of experiments aimed at deriving unknown, or at confirming assumed physical (or at least causal) principles behind the observed process, when outliers are basically related to the observation points or measurement values that cannot be explained with the known or assumed physical principles. This represents identification of outliers from a physical point of view, where outliers are values that are physically implausible, or in some other ways violate underlying physical principles assumed to be valid under the considered conditions (see also [25,51]).

It should be noted that analysing outliers from a statistical point of view is closely (but not exclusively) related to a concept of a visible outlier. Indeed, if statistical approaches are used for the identification of outliers, all identified outliers are by definition visible outliers, as they should be clearly distinguishable from the reference set of values to which they do not belong to. From the physical point of view, i.e., if underlying physical principles are used for the identification of outliers, the outliers may or may not be visible. Specifically, the outliers may be hidden in the following cases: (1) underlying physical principles are not fully and correctly describing the observed process; (2) the observed process is not properly measured; (3) the process is stochastic and features strong uncertainties; (4) the process is highly dynamic and an outlier is related to a recorded transient value;

(5) conditions simply cannot be properly reproduced or repeated in terms of both known and unknown parameters. Typically, a hidden outlier is found within the otherwise regular reference set of data and an example are (random) de-synchronised WS- P_{out} measurements, when some outliers may be recorded outside the reference cluster (i.e., they will be visible), while other may be inside the reference cluster (i.e., they will be hidden). In such cases, none of them should be used for the analysis, although hidden outliers will seemingly be in accordance with the underlying physical principles, making their identification much more difficult.

Accordingly, this paper uses a general but simple terminology, in which an outlier can have only two attributes: (a) true outlier, which means that this outlier should be removed (or suitably processed) from the further analysis, and (b) false outlier, meaning that this outlier could, or should be considered for the further analysis, assuming that there is a clear explanation or reasoning behind that decision, or that conditions under which this outlier is obtained are included in the model. An example are “mid-curve stacked outliers”, which might be due to wind curtailment control of a WT [29,39]. In that case, the decision whether they will be processed as the true or false outliers basically depends on the inclusion, or not, of curtailment mode of operation in the WT model [20,49].

This approach can be applied to outliers processed from both statistical and physical points of view, as well as for both visible and hidden outliers. Although not necessarily using the same terminology, a similar discussion about true outliers and false outliers can be found in a few references. For example, although many of the observed outliers are due to problems with the MET tower anemometers, there are some outliers that may not be errors, as individual WT sensors in general have measurements different from the MET tower measurements, which means they are false outliers [50]. “Isolated outliers” are one category of outliers discussed in [47], which may occur due to e.g., extreme weather conditions or due to human interference, and they may be either false (if extreme weather is to be included in the model), or true (if they are due to a human error). A clear distinction between the true outliers and false outliers is crucial in all wind energy applications, but especially in condition monitoring, where typically measurements from various sensors different from the measurements discussed in Section 2 are used. According to [13,25,27,30–35], false outliers may exist in the recorded data and may be very important for condition monitoring, but further analysis of these references is deemed out of scope of this paper, which focuses on the approaches for dealing with such outliers, not on the subsequent conditioning monitoring analysis.

3.2. Common Causes of Outliers in WT and WF Measurements

Reported causes and origins of outliers in processing WT and WF measurements are many and various. This section provides a general overview based on available literature, where some references provide a more detailed discussion of the origins in terms of the characteristics, manifestations and types, while some references also attempt to classify outliers based on their characteristics and to analyse their causes (these are discussed and further reviewed in Section 3.3).

In existing literature on processing WT measurements (WT-level analysis), the outliers may be consequences of:

- WT downtime, offline operation, or outage: [29]
- WT curtailment, down-rating, or derating: [20,29,31,39,49]
- (Manually set) WT constrained operation: [7,35]
- Data acquisition failure: [24,29,32,33,41,45]
- Inaccurate measurements caused by improper measuring location (usually for WS), or WS under-reading: [7,31]
- Data transmission (communication system) failure: [20,23,29,33,41,45,47]
- Data processing failure: [23,45]
- Data management failure: [20,23,33]
- Electromagnetic interference: [23,29]

- WT control system problem, including pitch control malfunction, blade pitch angle error, control program problem, yaw and pitch misalignment, incorrect or unsuitable controller setting and similar: [7,18,23,31,35,39,45,49]
- WT blade damage, or excessive deposition of dirt, ice or insects: [18,20,29,31,35,45]
- Wind shear: [50]
- Shading or wake effects from neighbouring WTs, terrain or physical obstacles: [7]
- Extreme weather conditions, harsh environment: [29,33,45,47,49]
- Air density fluctuations: [7]
- Issues related to averaging period: [7,41]
- WT cut-out effects: [41]
- WT data availability issues: [50]
- Other, including WT malfunctions, alarms in WT, low level of gearbox oil, worn-out generator brushes, sensor accuracy, and various errors in sensors: [18,20,23,24,32,35,45,47,49].

Regarding the processing of WF measurements (both directly and indirectly), the outliers may be caused by a number of similar sources as in the WT-level analysis. References listed below are explicitly discussing following sources of WF-level outliers (although they refer to problems, issues and conditions of individual WTs):

- WT downtime, offline operation, or outage: [48]
- WT curtailment, down-rating, or derating: [20,46,48]
- Data acquisition failure: [14,40,46]
- Data transmission (communication system) failure: [20,40,46,48]
- Data management failure: [20]
- WT blade damage, or excessive deposition of dirt, icing or insects: [20]
- Wind shear: [50]
- Extreme weather conditions, harsh environment: [40]
- Wind speed (around the cut-in or cut-out wind speeds): [14]
- WT data availability issues: [50]
- Other, including WT malfunctions, fluctuations in WT performance, multiple non-meteorological factors and various errors in sensors: [14,15,20,40,46,48]

3.2.1. Outliers Related to Data Acquisition, Data Transfer and Data Management Errors

The long-term measurements of WT and WF characteristics and parameters practically always result in large volumes of data [14]. Handling of these large volumes of data through the related procedures for data capturing, collecting, exchanging, processing, storing and using/managing is a particularly demanding task, which is a significant source of outliers. Accordingly, incorrect or corrupted data, or data errors, may occur in any of the four following general phases: (1) data acquisition; (2) data pre-processing and data processing (e.g., data compression, averaging, imputation of the missing data, etc.); (3) data transmission (data exporting/importing via communication systems); and (4) data management (e.g., selection and use of various data formats, data storage, data maintenance, etc.) [20,23,33,40,45,46].

If the outliers originate solely due to the errors and mishandling during the overall data acquisition, data exchange, data processing and data management procedures, they are defined as wrong measurements, data errors, incorrect data, corrupted data, or polluted/contaminated/dirty data. As they do not correspond to the actual measurement points, these outliers are true outliers, even when they are manifested as hidden outliers, i.e., when they occur within the range or boundary of the pre-specified regular/reference set(s) of data, or when they seemingly follow related underlying physical principles. As mentioned, their detection in such cases may be difficult, if not impossible, unless the information (e.g., from a separate log of events) is available for the periods of time when

there were sensor errors and other problems with measurement instruments, or when measured data were not properly processed, or when data were corrupted during the data transmission and data management procedures. Generally, if the outliers are not caused, or cannot be explained by the errors and mishandling in the data acquisition, transmission, processing and/or management procedures, these outliers will usually require further analysis and some of them are discussed in the further text.

3.2.2. Outliers Determined by the Operational Logs and Event-Logging Systems

Information on the presence of outliers can be simply obtained from the logs of operational events. In such cases, any related measurement, whether correct or incorrect, which is obtained during the period of outage, or testing, or maintenance of an individual WT is a true outlier and should be discarded from the analysis and not included in the operational model of the considered WT, as the conditions under which these measurements are obtained are not related to normal operating conditions (as indicated from the log). With the terminology proposed in this paper, however, these outliers are true outliers only at the individual WT-level. If the considered WT is part of a larger WF, and if the target model is operational PC WF model, all outage-related WT-level true outliers are part of the normal operational performance of the WF, what effectively makes them false outliers at the WF-level, i.e., these outliers should be included in the assessment of WF performance. For example, Ref. [50] reported that a WF with 274 WTs was operating with no more than 2 WTs out of operation (i.e., with 272, 273 or 274 WTs) for only 61% of time. Similarly, [58] reported that an on-shore WF with 6 WTs and an off-shore WF with 36 WTs were operating with all WTs only 30% and 14% of time, respectively, confirming that a common operational mode of WFs was with one or more WTs out of operation. In terms of the actual fault rates/probabilities of WT components, [59] states that these may vary from around 0.05 failure per year, to about 0.55 failure per year, with the repair times per failure varying from about 1.1 days, to 14.1 days.

Similarly, the information on the WT availability status is in [50] used as a separate log containing binary type of data, where 0 indicates unavailable WT and 1 indicates the WT is available. These input data are sampled and recorded at the start of the 10-min averaging window and [50] confirmed that these data also suffer from data quality issues: if the WT is producing power with an availability status of 0, there is a clear conflict. However, the opposite condition, when there is no power output from the WT, but the availability status is 1 (e.g., when the recorded WS is below/above the cut-in and cut-out WS's), is difficult to validate.

3.2.3. Outliers due to Applied Period of Averaging (Averaging Window)

Following the recommendations in [21], WS and P_{out} data are typically sampled at 1 Hz, i.e., obtained as second-by-second measurements, but are then averaged and recorded as the average 10-min data. It is interesting that [34] mentions that the purpose of using 10-min average values is to reduce the effects of outliers (as well as noise and auto-correlation). As WS typically exhibits strong short-term variations (in the range of seconds) and as (large) WTs usually have inertia time constants from a few to several seconds, this simply means that each pair of measured 10-min average WS and P_{out} values will typically consists of intra-10-min periods with both higher and lower than average recorded WS and P_{out} values. In the most general case, each 10-min period of averaging may include shorter periods of zero WS and P_{out} values (no wind, no production), when there will be also periods of (much) higher WS and P_{out} values (including wind speeds over cut-out WS, when WT will be stopped). For example, if the wind in the 10-min averaging window has intensive fluctuations, there may be consecutive WS recordings that are higher than the cut-out wind speed, triggering the control system to shut-down the WT, when the WT will re-start generation only if the WS decreases to a certain lower level. In some extreme cases, there may be multiple start-up and shut-down operations within a single 10-min averaging window and [41] explains that these outliers are not indications of faults or anomalous WT operation, basically sharing the same concept of false outliers in this review paper. These outliers are approached merely from a P_{out} -WS relationship in [7], where it is said that

P_{out} has a cubic dependency on the WS and therefore strong WS fluctuations in the averaging period will strongly influence measured P_{out} data.

3.2.4. Outliers due to Cut-Out Effects

The term cut-out effects in this section generally denotes the impact of the WT control system on the measured P_{out} values when the corresponding measured WS values are around and above cut-out wind speed. These measurements can be generally approached as outliers when the measured P_{out} values do not correspond to the values indicated by the Mfr-PC and are directly related to period of averaging and can be regarded as a special case of period averaging outliers from the previous sub-section. According to [41] and previous discussion, during a 10-min averaging window a WT may be stopped by the control system for several reasons, one of which is too high input WS, where the limit is typically denoted as a cut-out WS. If the WS within the averaging 10-min window is constantly (or at least mostly) higher than the cut-out WS, control will shut-down the WT and will keep it in that state, resulting in P_{out} (close) to zero, i.e., there will be no outliers in the recorded data. However, if the WS fluctuates around the cut-out WS, there might be periods within 10-min interval with WS lower than the cut-out WS, when WT will generate some output and when resulting 10-min average P_{out} values are somewhere between the rated power and zero, depending on the proportions of operational and non-operational times in the 10-min window. Essentially, similar averaging effects will also happen near the cut-in WS.

The common causes of outliers in WT and WF measurements are summarised in Table 2.

3.3. Characteristics and Causes of Outliers in Processing WT and WF Measurement Data

This section reviews characteristics and causes of outliers reported in available literature and illustrates them using representative P_{out} -WS scatter plots in Figures 1–3. Figure 1 illustrates WT-level outliers using actual 3-year 10-min synchronous nacelle WS and P_{out} measurements of a single WT, which are available to authors and are discussed in more detail in [3,29]. Figure 2a is based on the 3-year 10-min synchronous nacelle WS and P_{out} measurements of a single WT, which is pitch regulated, with cut-in, rated and cut-out wind speeds equal to 4 m/s, 18 m/s and 25 m/s, respectively [60]. Figure 2b plots the P_{out} -WS relationship of a whole WF, which consists of the 36 WTs of the same type as in Figure 2a. In that case, directly measured WF values of (single-sensor obtained) WS and (single-sensor obtained) P_{out} data were not available, so whole-WF values are calculated as the averages of synchronous WS and P_{out} measurements from all WTs in the WF [10]. Finally, Figure 3 shows WF-level data obtained from direct (single-sensor) synchronous 10-min measurements of WS and P_{out} values over the period of one calendar year, which are also available to authors.

According to a previously introduced distinction between the WT-level and WF-level analysis of outliers, discussion of causes and characteristics of outliers is divided in outliers that occur at both WT-level and WF-level, and outliers that will be highly unlikely to occur in one of these. It should be noted that if the processed WF measurements are obtained directly (single-sensor measurements of WS and single-sensor measurements of P_{out}), the analysis of these WF-level outliers is similar (but not identical) to the analysis and processing of outliers of individual WT measurements, i.e., WT-level outliers. In other words, some outliers at WT-level are highly unlikely to be found at WF-level. The greater differences from the WT-level outliers exist when WF-level data are obtained from the aggregation or averaging of measurements from individual WTs.

Table 2. Summary of common causes of outliers in WT and WF measurements.

Causes	Ref. (WT-Level)	Ref. (WF-Level)
WT downtime, offline operation, or outage	[29]	[48]
WT curtailment, down-rating, or derating	[20,29,31,39,49]	[20,46,48]
(Manually set) WT constrained operation	[7,35]	-
Data acquisition failure	[24,29,32,33,41,45]	[14,40,46]
Inaccurate measurements caused by improper measuring location (usually for WS), or WS under-reading	[48]	-
Data transmission (communication system) failure	[20,23,29,33,41,45,47]	[20,40,46,48]
Data processing failure	[23,45]	-
Data management failure	[20,23,33]	[20]
Electromagnetic interference	[23,29]	-
WT control system problem, including pitch control malfunction, blade pitch angle error, control program problem, yaw and pitch misalignment, incorrect or unsuitable controller setting and similar	[7,18,23,31,35,39,45,49]	-
WT blade damage, or excessive deposition of dirt, ice or insects	[18,20,29,31,35,45]	[20]
Wind shear	[50]	[50]
Shading or wake effects from neighbouring WTs, terrain or physical obstacles	[7]	-
Extreme weather conditions, harsh environment	[29,33,45,47,49]	[40]
Air density fluctuations	[7]	-
Issues related to averaging period	[7,41]	-
WT cut-out effects	[41]	-
Wind speed (around the cut-in or cut-out wind speeds)	-	[14]
WT data availability issues	[50]	[50]
Other, including WT malfunctions, alarms in WT, low level of gearbox oil, worn-out generator brushes, sensor accuracy, and various errors in sensors	[18,20,23,24,32,35,45,47,49]	-
Other, including WT malfunctions, fluctuations in WT performance, multiple non-meteorological factors and various errors in sensors	-	[14,15,20,40,46,48]

Figure 1. Actual 3-year measurements with indicated different types of outliers for an individual WT.

Figure 2. Actual 3-year measurements with indicated different types of outliers for: (a) an individual WT. and (b) a whole WF (derived from the individual WTs in the same WF).

Figure 3. Actual 1-year measurements with indicated constant WS-variable P_{out} outliers for a directly measured whole WF.

The characteristics and possible causes of various types of WT outliers are summarised as:

- Low P_{out} —High WS Outliers:** At WT-level, these outliers are characterised with zero, or very low, or even negative P_{out} values when the corresponding/synchronous WS values are between the cut-in and cut-out wind speeds, i.e., when WT is expected to generate non-negligible output power. These outliers are denoted as “bottom-curve stacked outliers” and “stacked data at the bottom of the curve”, typically characterised by horizontal dense data band in the PC-based model [29,39]. Their causes include WT failure or outage, unplanned WT maintenance, and faults of the WT measurement and/or communication systems [49]. Similarly, [24] stipulates that outliers representing P_{out} data close to, lower than, or equal to zero are exclusively due to wrong measurements or WT malfunctions (see Figure 2a and outliers marked with A). If the WT does not rotate, P_{out} should be zero, but if the WT’s control system is kept energised, the P_{out} measurements might be negative [26]. At WF-level, when data are obtained directly, these outliers are most likely true outliers due to measurement errors or communication system errors [48], or disconnection of the whole WF due to activation of the protection system (see Figure 1, outliers marked with ①). When WF-level data are obtained from the aggregation of individual WT measurements, these outliers are highly unlikely to be present (see Figure 2b, outliers marked with A). Regardless of WT-level or WF-level analysis, some of these outliers may be hidden, e.g., for low WS values.
- High P_{out} —Low WS Outliers:** At WT-level, these outliers are typically manifested as the horizontally stacked data with a narrow range of variations of relatively high P_{out} values (e.g., close to 1pu), which are clearly visible when they are above the upper boundary of usual/expected/regular reference WT PC values, i.e., for given WS values, these outliers represent higher than expected and relatively constant P_{out} values. These outliers are denoted as “top-curve stacked outliers” and “stacked data at the top of the curve”, which appear to be one or more horizontal dense data bands,

located above and to the left of the regular PC-based WT model (see Figure 1, outliers marked with ③) [29,39]. These outliers are usually caused by communication errors, or wind speed sensor failures ([29] states that faults of WS sensors happen frequently), e.g., when lower than actual WS values are recorded. When WF-level data are obtained directly, these outliers are similar to WT-level outliers [40,48], but may have different P_{out} values, based on the number of operational WTs in the WF. When WF-level data are obtained from the aggregation of individual WT measurements, these outliers may or may not be visible, based on the operating point of the WT(s) with faulted measurement/communication system, as well as the (aggregated) operational point of the whole WF.

- Low P_{out_max} Outliers: These outliers are again represented by horizontal bands/ranges of relatively constant P_{out} values, which are below the lower boundary of the expected range of PC values (see outliers marked as B in Figure 2a,b). These outliers are denoted as “mid-curve stacked outliers” and “stacked data in the middle of the curve” [29,39]. Mentioned causes include curtailment, but also down-rating of WTs and data acquisition and communication system errors [20,49]. In some cases, improper WT operation, for example due to a damage of WT gearbox bearing [27], may also limit the maximum WT P_{out} (e.g., restricted to 60% of the rated maximum power). When WF-level data are obtained directly, these outliers are similar to WT-level outliers [48] (see outliers marked by ② in Figure 1), but may have different P_{out_max} values, based on the number of operational or curtailed WTs in the WF ([50]) and may also include communication and measurement errors of single P_{out} and WS sensors [48]. For aggregated WF-level data, measurement and communication errors of individual WTs will most likely make these outliers to be hidden, but there may be many horizontal bands, based on the actual number of WTs that are curtailed or down-rated. An important feature of this type of outliers, distinguishing them from the next discussed type, is that there are no big differences in WF power outputs up to the point of curtailment. In terms of true/false outlier analysis, if curtailment (or downrated/derated) operation should be included in the WT/WF model, these outliers are false outliers.
- Shifted PC Outliers: At WF-level, these outliers are not consequence of curtailment or down-rating of individual WTs, but are due to the outages or faults of individual WTs in a WF, reflected in shifting of the PC to the right, i.e., corresponding to the situations in which both maximum power output is reduced and higher WS's are required to produce the P_{out} values close to these when all WTs are in operation (see outliers marked as C in Figure 2b). A WT outage is defined as the tripping of n WTs in [48], for e.g., protection reasons or unplanned maintenance. Similar outliers may be recorded for an individual WT, when PC is shifted to the left, indicating improper operation, or problem with WT control system, or damaged WT, or inaccurate speed sensor readings, as illustrated in [31].
- Linear P_{out} —WS Outliers: These WT-level outliers appear as linearly related P_{out} -WS recordings, possibly occurring during the data processing phase, when linear interpolation is applied to populate missing recordings (see outliers marked as C in Figure 2a). Similar outliers are reported due to malfunction of pitch-control system and dirt deposits on the blades [31]. These outliers may occur at WF-level when data are obtained directly (data recording errors), but they are highly unlikely to be visible when WF-level data are obtained from the aggregation of a larger number of individual WT measurements.
- Scattered Outliers: These outliers are related to irregular or random values around the expected/reference PC range (Figure 1, marked as ④), which may be due to faults and errors, but also due to statistical processing and averaging window. These outliers are called “around-curve outliers” in [29] (in [39], “scattered data around the curve”) and can be caused by random factors, like signal propagation noise, sensor failure and extreme weather conditions. Alternative term is “sparse outliers” (due to random noise) [20]. Also, [37] shows these outliers together with other types of outliers in WF measurements. The reasons for these abnormal data are sensor failure, sensor noise and some uncontrolled random factors [49]. At WT-level and when WF-level data

are obtained directly, these outliers have similar causes. When WF-level data are obtained from the aggregation of individual WT measurements, their dominant cause are averaging window and similar statistical processing-based origins (see Section 3.2.3). These outliers are termed as “unnatural data” [40], where they are analysed separately from other outliers (these sparsely located data can also be seen in Figure 3).

- **Constant WS-Variable P_{out} Outliers:** These outliers are reported at WF level and are manifested as vertical bands of constant WS values (or only with small WS variations) from the MET towers recorded over a longer period with synchronously recorded relatively large P_{out} variations, occurring due to errors in data acquisition and data transmission (polluted data) [40,50] (see Figure 3). These constant WS outliers are more strongly pronounced at WT-level and when WF-level data are obtained directly, but may be less pronounced when WF-level data are obtained from the aggregation of individual WT measurements. A variant of the constant outliers is denoted as “slender” in [40], related to approximately vertical band around the cut-out wind speed, explained by the wake effects, i.e., that the WTs within a WF do not cut out together near the cut-out wind speed, because the wind speeds at each WT vary from the measured WS at the MET mast [40]. It is emphasised that these data should be categorised as valid data, as they reflect the natural P_{out} fluctuation property of the WF around the cut-out WS, which is important to the system operators. However, it is not clear why these outliers are not scattered within an oblique band with a negative slope (as shown in Figure 2b) for outliers marked with E), but are clustered in a relatively narrow vertical band near the cut-out wind speed (wake effect outliers as discussed in [40]). When WF-level single-sensor measured WS is around the cut-out WS, most likely situation is that some of the WTs in the WF will be operating (their WS is below the cut-out wind speed), while some WTs will be stopped (as their WS is above the cut-out wind speed) and therefore the 10-min average P_{out} values for the whole WF will be between rated power and zero when the single-sensor measured WS's are around cut-out WS. According to [40], it is necessary to distinguish “(invalid) unnatural” outliers with these “valid data”, but this can be a difficult task, since both types of data may be caused by WT cut-out effects. The methods and results of [40] are further discussed and analysed in Sections 4 and 5.

The characteristics and causes of outliers in WT and WF measurements discussed above are summarised in Table 3.

Table 3. Summary of characteristics and causes of outliers in WT and WF measurements.

Outlier Characteristics	Possible Causes	Ref. (WT-Level)	Ref. (WF-Level)
Low P_{out} -high WS	WT failure or outage, unplanned WT maintenance, or faults of measurement/communication system	[24,26,29,39,49]	[48]
High P_{out} -low WS	Measurement/communication errors, or wind speed sensor failures	[29,39]	[40,48]
Low P_{out_max}	Curtailment, damage of WT gearbox bearing, or measurement/communication system errors	[20,27,29,39,49]	[48,50]
Shifted PC	Outages (caused by protection or unplanned maintenance) or faults of individual WTs in a WF	-	[48]
Linear P_{out} -WS	Linear interpolation, malfunction of pitch-control system, or dirt deposits on the WT blades	[31]	-
Scattered	Signal propagation noise, sensor failure, extreme weather conditions, averaging window, or uncontrolled random factors	[20,29,39,49]	[37,40]
Constant WS-variable P_{out}	Errors in data acquisition and data transmission, wake effects, or P_{out} fluctuation around cut-out	-	[40,50]

3.4. Data Rejection Requirements in IEC Standard 61400-12-1

The IEC Standard [21] does not provide detailed or specific discussions of outliers, but it sets rules for data rejection, i.e., it specifies which data shall be excluded from the analysis, in order to ensure that only data obtained during normal operating conditions of the WT are used in the analysis (and to ensure that used data are not corrupted). Accordingly, [21] states that during the measurement period the WT shall be in normal operation, as prescribed in the WT operations manual. Also, the machine/WT configuration shall not be changed.

According to [21], the data shall be removed under the following circumstances (other applied rejection criteria shall be clearly reported):

- When external conditions, other than wind speed, are out of the specified WT operating range;
- If WT cannot operate because of fault condition;
- When WT is manually shut down, or it is in a test, or maintenance operating modes;
- If there is a failure or degradation (e.g., due to icing) of measurement equipment;
- When WD is outside the measurement sector(s), which generally exclude WDs with significant obstacles and other wind turbines, as seen from both the WT and measurement equipment;
- When WDs are outside the valid (complete) site calibration sectors;
- For any special atmospheric condition filtered during the site calibration, which shall also be filtered during the power curve test.

In particular, [21] emphasises that a large hysteresis loop in the WT cut-out control algorithm may have considerable effects and therefore shall not be included in the obtained WT PC model. It further instructs that all data sets where the WT has stopped generating power due to the cut-out at high wind speed shall be excluded. This is difficult to understand and to acknowledge, as the WTs are typically installed at locations with good wind energy resources, i.e., at locations where it is likely that WS will reach and go above the cut-out WS, so these operational conditions and their considerable effects should not be excluded from the analysis. However, [21] suggests to collect and present measurements with cut-out behaviour in a special database and further comments that neglecting cut-out effects may lead to overestimations of energy production, especially for cases with higher average WS's. On the other hand, [21] treats cut-in effects differently, saying that power curve model shall capture the effect of hysteresis at the cut-in control algorithm, as well as the effect of parasitic losses below cut-in WS, which is possibly due to a much larger number of measurements around cut-in WS, than around cut-out WS.

Importantly, [21] suggests that all subsets of data collected under special operational conditions (e.g., high blade roughness due to dust, salt, insects and ice deposits, or if grid conditions vary significantly), or atmospheric conditions (e.g., precipitation, wind shear) that occur during the measurement period may be selected as special databases, which basically means that these values are treated as false outliers (see [31] for further discussions and illustrations of these outliers).

Finally, it should be noted that the rejection of data will result in missing values and these enforced missing data should be distinguished from the missing measurements or recordings, e.g., due to the failures of the data acquisition/measurement systems.

4. Detection of Outliers in Processing Wind Turbine and Wind Farm Measurements

To the best of authors' knowledge, there is a lack of a comprehensive and systematic research on the detection and treatment of outliers in wind energy applications, [29], as opposed to some other areas, such as biological data, astronomy data, web data, information network data, economics data, and medical data, etc., where comprehensive and structured reviews are available (e.g., [61,62]). Specifically, according to [62], approaches for detection of outliers can be divided in three general groups: unsupervised, semi-supervised and supervised, depending on the availability of data labels, i.e., availability of information on conditions under which the data are obtained. Unlike the

literature related to other areas, which usually adopt general statistical approaches, one important contribution of this review paper is that it specifically focuses on the WT and WF applications, providing inputs, illustrations and specific analysis for understanding the causes of outliers, aimed at a better understanding of WT and WF operations through establishing robust methods for detecting true outliers. In terms of WT-level and WF-level measurements, there is usually a lack of documentation and information about important operational conditions, such as wind curtailment [20,29,48], WT outage [48] and WF operating states [40]. Similarly, information on the possible causes of outliers, such as data storage mishandling and data communication errors, or WT damage or presence of icing on blades, may also not be documented [20,48]. As discussed in Section 3.2.2, information from operational and event logs may significantly improve detection of outliers [40]. In particular, [29] considers wind curtailment as the source of abnormal data simply because information on wind curtailed operation may not be available. On the other hand, [48] states that planned WT maintenance is not considered as a source of outliers, because the planned maintenance is usually well documented.

The logs with the information on WT downtimes, malfunctions, faults and maintenance periods can be used, but these raw data are not free from errors and may also require data cleaning, as [25] states. However, this reference focuses on condition monitoring, rather than resolving data quality problems, and fault labelled data sets are actually used for WT abnormality detection. Finally, [46] describes a semi-supervised algorithm for outlier detection, but it uses a limited number of labelled data.

Most of the outlier detection methods in reviewed literature belong to unsupervised approaches, as they did not have, or did not use any further explicit information on the exact causes of outliers. Like in previous discussion, the reported outlier detection approaches reviewed in this section can be also divided in statistical methods, physical constraint-based methods, or combinations of both. Robust statistical approaches proposed in some papers are also reviewed, although this work is not related to outlier detection, but to reducing the effects of outliers.

4.1. Statistical Methods for Outlier Detection

Statistical methods for the identification of outliers are most common and can be divided in density-based (aimed at separating sparse and dispersed points from the highly concentrated ones), distance-based (aimed at isolating points located far away from the normal ones, which usually can also be regarded as equivalent to the density-based method), correlation-based (aimed at finding the outliers using joint or conditional correlation analysis), and image-based (aimed at detecting the abnormal data by extracting the principle part from the binary image of P_{out} -WS scatter plot).

Regarding statistical methods, the most popular outlier detection algorithm is quartile algorithm, which is usually combined with other more complex approaches. In descriptive statistics, the interquartile range (IQR) is defined as $IQR = Q_3 - Q_1$, where Q_3 and Q_1 are 75th and 25th percentiles, respectively. Quartile algorithms for outlier detection usually demonstrate that any data point outside the range $[Q_1 - k \times IQR, Q_3 + k \times IQR]$ may be an outlier, where specifically, $k = 1.5$ indicates outliers and $k = 3$ indicates data that are “far out” [63].

Another frequently used approach for outlier detection is standard deviation-based method, which usually assumes that the analysed data have normal distributions. However, according to [51] and [64], the normality assumptions usually do not hold, as the P_{out} data in any WS bin cannot pass conventional statistical tests, e.g., Kolmogorov–Smirnov test and Lilliefors test. Like quartile algorithm, this method is usually applied with other more advanced approaches. According to the standard deviation-based method, for a set of sample data $x_1, x_2, \dots, x_n$, any point outside $[\bar{x} - k \times \sigma, \bar{x} + k \times \sigma]$ may be selected as an outlier, where $\bar{x}$ is the mean value of samples, σ is the standard deviation of samples and k is used to set the acceptable range of deviations.

4.1.1. Wind Turbine Measurements

Instead of applying quartile algorithm directly on the measurements, [25] argues that the negative deviations in some WS bins should be prevented and it applies sideband normalisation

to regularise the deviations of the graded statistical features before applying the quartile algorithm. In [29], quartile algorithm is combined with a change detection approach (i.e., change point grouping algorithm) to detect the outliers, where change point is defined as a point where one or more quantities in a series exhibit sudden change, therefore reflecting a qualitative change in the data. In terms of the stacked outliers (see outliers in Figure 1 marked as ①, ②, ③ and outliers marked as A and B in Figure 2a), the distribution characteristics of the large amount and highly concentrated abnormal data imply that the mean value, the variance and the change rate of related data sequence change as well [65]. In this reference, WS data are divided into several bins and the variance change rates of the ordered P_{out} values in each bin are calculated, in order to find the change points for distinguishing outliers from the normal data. As the change point grouping algorithm cannot efficiently identify all types of outliers, especially the scattered outliers (e.g., outliers marked as ④ in Figure 1 and marked as D in Figure 2a), [29] also adopts quartile algorithm to perform further outlier detection.

Unsupervised clustering algorithms are another widely applied method for analysing outliers in WT measurements, as [20] shows. This reference uses density-based spatial clustering of applications with noise (DBSCAN) algorithm [66], where main idea is that for each point of a regular data cluster, the neighbourhood of a given radius must contain at least a minimum number of points. Accordingly, the approach in [20] first identifies boundary of outliers, with the purpose of eliminating stacked outliers, and then applies quartile algorithm twice to identify sparse outliers, equivalent to detecting the outliers for both $Prob(P_{out}|WS \text{ bin})$ and $Prob(WS|P_{out} \text{ bin})$, which are notations of conditional probabilities: $Prob(P_{out}|WS \text{ bin})$ is the conditional probability of P_{out} given WS bin, while $Prob(WS|P_{out} \text{ bin})$ is the conditional probability of WS given P_{out} bin.

Similarly, [33] applies a sliding window technique to divide data into different segments that are considered as different objects, where their attributes consist of time-domain statistical features (e.g., mean, maximum and peak-to-peak values) extracted from each segment. Then, an improved local outlier factor (LOF) algorithm, known as kernel-based LOF (KLOF) algorithm [67], uses these attributes to evaluate whether the segments contain outliers. Traditional LOF algorithm [68] is a density-based unsupervised learning approach, which detects abnormal data by measuring their local deviations from the neighbours, basically sharing some concepts with DBSCAN algorithm, like core distance and reachability distance. Traditional LOF algorithm has two main shortcomings: (1) it may fail to detect outliers in a complex and large set of data, because it calculates the averages of reachability distance between an object and its neighbours in the local density estimation; and (2) it is very sensitive to the parameter selections. The KLOF algorithm transfers the unsupervised learning into the classic non-parametric regression learning and it tries to overcome the above two shortcomings of LOF by computing the regression estimator of a data point in its neighbours based on Nadaraya-Watson kernel regression.

In addition to DBSCAN, LOF and related improved algorithms, there are other clustering methods used in reviewed literature, as [45] demonstrates. This reference applies k -means clustering algorithm based on Squared Euclidean and City-Block distance to categorise the raw P_{out} -WS data into different groups. The detections of abnormal data are performed in the created clusters, where the local Mahalanobis distance of each data point to its cluster centroid is used to detect the outliers based on their score. Similarly, Ref. [15] also uses k -means clustering algorithm with a conservative approach based on the Mahalanobis distance metric to identify outliers. The main difficulties in traditional k -means clustering algorithms are selection of appropriate number of clusters, k , and sensitivity k -means of obtained results to the initialisations. Therefore, Ref. [39] uses an improved version of k -means clustering algorithm [69], where residual analysis method is used to automatically obtain both the number of clustering and initial centres from the decision diagram. Furthermore, Ref. [47] proposes an unsupervised outlier detection approach combining stacked denoising autoencoder (SDAE) [70] and density-grid-based clustering method [71,72], where SDAE is utilised to extract data features, followed by the density-grid-based clustering method to detect outliers.

Some of the previous work used classification methods for outlier detection. One example is [35], in which k -nearest neighbour (KNN) classification algorithm [73] is used first to predict the desirable power output for the WT operating in normal conditions, which is constructed as a reference WT PC, and then outliers are detected using the residual approach (control theory) and control charts (quality control) [74] on the established PC model. In that way, KNN algorithm is used as a supervised classification algorithm, which is trained on a selected (training) set of samples in a multidimensional feature space, where the labels of the training samples must be known. In the classification phase, for an unlabelled vector (a univariable or high-dimensional sample point), the algorithm will calculate and compare the distances from the k training samples, in order to determine the label to assign. Unfortunately, [35] does not illustrate how the labels for the training set are obtained.

In some references, correlation-based outlier detection models are used, where outliers are detected using Copula theory [75]. More specifically, Ref. [41] applies a 2-dimensional Gaussian mixture copula model (GMCM) [76] to fit the P_{out} -WS joint distribution, i.e., $\text{Prob}(P_{\text{out}}, \text{WS})$, where the outliers are excluded by a joint probability contour. Similarly, Ref. [42] also uses outlier detection method based on Copula. It uses mixed Archimedes Copula function to calculate the conditional distribution of WT P_{out} for given WS, i.e., $\text{Prob}(P_{\text{out}}|\text{WS})$, from which the confidence boundary of P_{out} is obtained. Acknowledging that the operation characteristics of a WT may change over time, this reference uses k -cross validation (k -CV) and sliding time window (STW) approaches to determine whether the estimation is stable.

Normal-distribution-based outlier detection approaches are also often reported in previous literatures. Reference [43] first applies exponential smoothing [77] on WT P_{out} series to account for WS fluctuations, where the original P_{out} series is denoted as $x_1, x_2, \dots, x_n$, and the transformed series as $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n$. Afterwards, outliers are detected by a standard deviation-based approach, as data points outside the range $\bar{x}_{t-1} - k \times \sigma < x_t < \bar{x}_{t-1} + k \times \sigma$, where the subscript t denotes the data point at the t -th step, σ is the standard deviation, and k is a parameter determined from statistical analysis of small probability events ($k = 3$ is the choice in [43]). Similarly, Refs. [18,24] both assume that distributions of WT P_{out} values in WS intervals/bins follow normal distributions and then use multiple standard deviations around the mean values to determine the outliers. Furthermore, Ref. [26] performs recursive data cleaning by utilising the piece-wise linear model as a benchmark to remove the outliers, where the main criterion is that for a large number of data samples, there should be at least a certain percentage of normal observations distributed within ρ standard deviations from their associated fitted line segments. Finally, Ref. [23] proposes a deterministic PC model, considering both WS and WD dependencies, i.e., $f(P_{\text{out}}|\text{WS}, \text{WD})$ using least median of squares (LMedS) fitting algorithm. One significant advantage of LMedS, compared to least mean square fitting algorithm, is that it is more robust as the fitting results are determined by the median values instead of mean values, as the outliers are detected as the points with high residues according to the fitted model $f(P_{\text{out}}|\text{WS}, \text{WD})$, assuming the noise for the inliers has a normal distribution. The standard deviations of the errors from LMedS residuals can be estimated, which determines the thresholds to distinguish outliers from the inliers.

As discussed above, many outlier detection approaches are aimed at identifying the boundaries of possible P_{out} -WS ranges, and [31] develops a unique automatic PC limits calculation algorithm through iterations. The presented algorithm moves the estimated PC left/right or up/down while it searches for the optimal PC limits, trying to include only normal data. In the iteration loop, only the identified normal data from the previous iteration step are selected in each step as the new input data. The iterations stop when the percentage change of identified normal data is below a certain level (this approach is also used in [34]).

Examples of using hypothesis test theory to identify outliers are also found, where e.g., Ref. [44] uses a combination of four methods to determine whether a data point is an outlier, representing a cyber-attack state in a system state vector space. Firstly, an F-test [78] is employed, and the null hypothesis is that the data in two vectors come from the same variance for an attacked state vector.

Because a small number of outliers will not lead to significant variance change, a majority voting is used if there is no sufficient evidence to reject the null hypothesis. If the results of at least two of three algorithms: fuzzy c-means (FCM) clustering algorithm [79], quartile algorithm and median absolute deviation (MAD) method [80], show that a data point is an outlier, then that data point is recognised as an abnormal data point.

A relatively new technique in the WT measurements outlier detection research is presented in [49], based on WT PC images. After pre-cleaning of some obvious and easy-to-detected outliers (e.g., ① in Figure 1 and A in Figure 2a), the obtained WT P_{out} -WS scatter plot is transformed, so it will form a corresponding binary image. Then, the principal part of the binary image, representing the normal data, is extracted by the mathematical morphology operation (MMO) [81], which is combined with the Hu moment [82].

4.1.2. Wind Farm Measurements

Regarding the WF measurements, [46] uses a limited number of labelled data to supervise a DBSCAN algorithm for the detection of outliers in WF measurements in order to avoid false dismissals and false alarms. The limited number of labelled data is used to form a semi-supervised approach that could lead to a good balance between artificial label-setting of supervised learning and detection precision, despite the fact that DBSCAN itself is an unsupervised technique. Reference [40] utilises LOF algorithm using weighted Euclidean distance as a similarity measurement to detect outliers in WF measurements. In particular, as discussed in Section 3.3, [40] states that the vertical band around the cut-out wind speed in the WF P_{out} -WS scatter plot are not outliers, but are caused by wake effects. The results presented there show that LOF algorithm also does not classify these recordings as outliers, which could be simply due to the fact that these data have high densities, as otherwise LOF, a density-based clustering algorithm, will not be able to denote them as the normal data.

The KNN classification based algorithm is also used to detect outliers in WF measurements, with [14] applying approach very similar to [35], as discussed in Section 4.1.1. Correlation-based outlier detection models are applied for WF measurements as well. Copula theory is used to fit the correlation model in [48], where probabilistic WF PC is obtained by the Copula conditional quantile method, $Prob(P_{out}|WS)$, and where outliers are identified as the points outside its upper and lower boundaries. Finally, Ref. [37] first uses a modified moving average algorithm to smooth the raw WF measurements and then applies following censoring rule to remove the abnormal data: if $y_i > a_1(x_i^3)^2$ or $y_i < a_2x_i^3 \cap y_i < a_3$ ($i = 1, 2, \dots, n$), then i -th element is removed, where x is the WS, y is the normalised P_{out} measurement, and a_1 , a_2 , and a_3 are the parameters that should be tuned.

4.2. Physical Constraint-Based Methods for Outlier Detection

There is a relatively small number of references dealing with the detection of outliers based on physical constraints. Furthermore, most of the rules applied for setting the constraints are ambiguous. For example, [7,27] state that inputs and outputs of the WT measurements must be within the expected ranges, which have to be chosen extremely carefully, and that inputs and outputs should be mutually consistent. However, these references do not show explicitly how the ranges and consistencies could be determined, as they focus on conditioning monitoring, not on data pre-processing for the identification of outliers. Regarding the WF measurements, [50] states that the first step is to remove recordings when less than 272 WTs are available in a WF with total of 274 WTs, adding that this is an arbitrary criterion, so that the remaining data will have less than 1% uncertainty in P_{out} , since $2/274 = 0.73\% < 1\%$. Afterwards, [50] performs general outlier detection checks, removing outliers when: (1) positive P_{out} values occur for $WS < 2$ m/s; and (2) constant WS values are recorded for longer than two consecutive 10-min averaging windows.

4.3. Outlier Detection by Combinations of Statistical and Physical Constraints-Based Methods

Reference [51] firstly determines the upper bound of possible WT P_{out} values according to Betz's law, which specifies the maximum power output for given air density, rotor area and input wind speed. Then, it uses method of bins and quartile algorithm to detect the remaining outliers, stating that if the WT P_{out} histogram has multiple modals, only the highest populated interval in the histogram is considered, while P_{out} recordings near the lowest mode are excluded.

Table 4 summarises possible problems or limitations of outlier detection methods reviewed in Sections 4.1–4.3.

4.4. Robust Statistical Models for Reducing Outlier Effects

Robust statistical models are proposed in some references to reduce the effects of outliers, where, arguably, the main advantage is that an explicit detection of potential outliers is not performed. For example, hidden Markov model (HMM) based on fuzzy scalar quantisation [83] is proposed in [30] for WT fault diagnostics. As few outliers occurring in the HMM training set may lead to a misdiagnosis, fuzzy scalar quantisation assigns to each state in the signal quantisation a discrete probability, rather than a single value, to reduce the effects of such outliers. Furthermore, [32] employs the weighted version of least squares support vector regression (weighted LS-SVR) from [84] to model the relationship between weather conditions and WT P_{out} , stating that a single additional weighted LS-SVR iteration is generally sufficient for excluding the outlier effects by giving minimal weights for the observations with large errors following specific rules [85]. Finally, in [52] WF P_{out} prediction is analysed by combining outlier smooth transition autoregressive (OSTAR) approach and generalised autoregressive conditional heteroskedasticity (GARCH) models, where resulting OSTAR-GARCH model is used to capture the regime switching between the “outlier shocks” and other values in the volatility of WF P_{out} time series, where outlier shocks are defined as the values outside $\pm 2\hat{\sigma}$ interval, with $\hat{\sigma}$ being the estimate of the residuals' standard deviation.

Table 4. Summary of possible problems or limitations of reviewed outlier detection methods in practical applications.

Method		Possible Problems or Limitations	Refs.
Type	Name		
Classification	KNN	Plenty of normal data as samples; computational burden	[14,35]
Clustering	<i>k</i> -means	Usually difficult to determine the parameter <i>k</i>	[15,45]
	Improved <i>k</i> -means	Does not work well for abundant outliers	[39]
	DBSCAN	Sensitive to manually set parameters; does not work well for abundant outliers	[20]
	(Semi-supervised) DBSCAN	Limited number of labelled data may not have the same statistical properties as the test data	[46]
	LOF	Does not work for stacked outliers; does not work well for abundant outliers	[40]
	KLOF	Does not work well for abundant outliers	[33]
	SDAE and density-grid-based clustering	Does not work well for abundant outliers	[47]
	FCM	Does not work well for abundant outliers	[44]
Correlation	Copula	Stacked outliers could not be effectively filtered	[41,42,48]
Distance	Normal distribution	Assumption of normal distribution is not always suitable [42]	[18,24,43]
	Quartile algorithm	It is not effective when the proportion of outliers is large	[25,29,51]
	MAD	The selection of MAD range may not be applicable in specific situations [80]	[44]
Change point	Change point grouping	Need to manually set parameters; ignores the overall distribution; does not work well for the abundant stacked outliers	[29]
Recursive or iteration data removing		Only work for data set with large number of recordings, so the algorithms can converge; difficult to determine a stopping criterion	[26,31,34]
Statistically robust data fitting		Computational burden [23]	[23]
Data smoothing and censoring		Parameters must be carefully tuned to adapt to different cases [20]	[37]
Computer visual		Cannot distinguish false outliers near the cut-out wind speed	[49]
Hypothesis test	F-test	Assumes that data are normally distributed and that samples are independent [78]	[44]
Physical	Various consecutive WS	Cannot detect outliers caused by P_{out}	[50]
	Expected ranges and mutually consistent	Cannot prove the correctness of selected ranges	[7,27]
	Betz's law	Can only determine the theoretical upper PC bound	[51]

5. Approaches for Testing Success of Outlier Detection and Removal/Treatment Procedures

There is a general lack of benchmark data and labelled, or well-documented data that can be used for testing the accuracy/success of outlier detection and removal/treatment methods. The most widely used method to evaluate the success of outlier detection/treatment approaches is a simple visual check. For example, [13] uses “visual inspection” to ensure that resulting power curve is without outliers, while [26] compares the scatter plots before and after removing outliers concluding: “... it can be seen that the outliers included in the raw data are effectively eliminated in the final data set”.

In some literature, outlier elimination rate is used as an indicator of the performance of proposed outlier detection/elimination methods. Other widely employed metrics include: confusion matrix, sensitivity (i.e., true positive rate, TP rate), specificity (i.e., true negative rate, TN rate), false positive rate (FP rate), false negative rate (FN rate), and F-score (i.e., F-measure) (e.g., [23,44,46–48]). Each row of confusion matrix represents the instance in an actual class, while each column represents the predicted instance, or vice versa. TP and TN are outcomes where the model correctly predicts the positive and negative classes, respectively. FP means the model predicts the positive class by mistake, while FN refers to the model mistakenly predicting the negative class. The formulas of TP rate, TN rate, FP rate, FN rate and F-score are listed as following:

$$\text{TP rate} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (1)$$

$$\text{TN rate} = \frac{\text{TN}}{\text{TN} + \text{FP}}, \quad (2)$$

$$\text{FP rate} = \frac{\text{FP}}{\text{FP} + \text{TN}}, \quad (3)$$

$$\text{FN rate} = \frac{\text{FN}}{\text{FN} + \text{TP}}, \quad (4)$$

$$\text{F-score} = \frac{\text{TP}}{\text{TP} + 0.5 \times (\text{FP} + \text{FN})}, \quad (5)$$

Furthermore, the comparisons of the statistical properties between the measurements and known data sets without outliers are also considered (e.g., [48]), while comparisons of power curve fitting errors and comparisons of data concentration indices are used to prove the correctness of outlier removing approaches. More specifically, Ref. [29] emphasises that the outliers can be effectively identified: “the data deleted in the three turbines are basically outside the normal power output ranges, indicating that this outlier identification algorithm is effective”, as the remaining normal data in the WT P_{out} -WS scatter plot are “close to the ideal-state wind power curve”. However, the definition of normal power curve range is not provided in this reference, which also compares the proposed outlier identification method with other methods based on the outlier elimination rate. Furthermore, Refs. [20,39] both calculate the outlier elimination rates and compare PC modelling errors before and after removing outliers from the raw measurements, stating that “for a well performed outlier elimination method, both errors of the power curve modelling and the elimination rate should be as small as possible”. Additionally, Ref. [40] states: “the simplest way to evaluate the accuracy of the algorithm is by visual inspection” and uses confusion matrix to show the effects of the outlier detection method, but it does not illustrate how the true states in the confusion matrix are obtained. Finally, Ref. [42] uses R-squared method to indicate the concentration of data, declaring that a higher concentration means better data cleaning effect.

In terms of a strict analytical or mathematical formulation, visual check is not an accurate and scientific way to assess the results of outlier detection methods. The use of outlier elimination rate as a performance index is also unreliable, especially when there is a lack of labelled data, or otherwise clearly indicated outlier data. Essentially, obtaining both higher and lower outlier elimination rate is misleading, because either does not imply that one method is better or worse than another. Furthermore,

comparison of power curve fitting errors, or data concentration indices, may be meaningless, as almost all outlier removing methods lead to a filtered data set with higher concentration, which will certainly have better fitting results and smaller residuals. However, evaluations through confusion matrix or comparisons of data statistical properties are reasonable, but they inevitably require knowledge of the true outlier labels in the measured data.

In terms of the treatments of the identified outliers, the majority of previous work simply decide to eliminate (i.e., to remove, or delete, or filter-out) outliers after they are detected, without any further processing, e.g., [7,14,15,18,20,23,24,26,27,29,31,34,35,37,40,41,43,45,50,51]. For example, [41] excludes sparse outliers caused by window-averaging and WT cut-out effects (as discussed in Sections 3.2.3 and 3.2.4), although it states that these values are not due to the faults or otherwise anomalous WT operations. Similarly, [41] keeps the outliers caused by WT curtailments, which is probably because of the used outlier detection method, GMCM, in which these outliers will have higher joint probability densities than the threshold, allowing to identify them as normal data.

Although not explicitly recovering deleted outliers, [20] mentions that these recordings can be corrected by temporal and spatial interpolation methods [86–88], and that the fitted power curve based on normal measurements could also be utilised to infer the P_{out} values for given wind speeds as inputs. Furthermore, [40] states that some imputation approaches (e.g., [89]) can be used to replace (or predict) missing data due to outlier filtering, but it argues that these efforts may not be necessary for two reasons: (1) the outliers are often consistent and lasting for longer periods, and there are not sufficient data to make a smooth imputation; (2) there will be enough remaining data to obtain interesting patterns, as usually a relatively small portion of outliers is removed. Similarly, Ref. [27] also stipulates that neglecting outliers instead of approximating them is acceptable, since there is a large number of remaining data available after the outliers are neglected. However, that may not be true in case of e.g., measurements around and above cut-out WS, which are generally with low probability (e.g., in a WT example in Figure 2a, only 0.5% of the measurements over a 3-year period were recorded for WS's around and above the cut-out wind speed).

A very few references in existing literature try to recover outliers by data imputation. An example is [42], which proposes a bi-directional Markov chain interpolation method to recover the missing values, stating that there will inevitably be a lot of missing data if the stacked outliers caused by wind curtailment are eliminated. Another example is [53], where the outliers are treated as missing values and imputation of these data are performed using multilayer perceptron (MLP) and adaptive neuro-fuzzy inference (ANFIS) networks [90,91], stating that it is crucial to have an uninterrupted historical data set, especially for forecasting applications. Finally, [48] exploits the spatial correlation between adjacent WFs to correct the P_{out} outliers in a WF, but outliers in the WS data are not considered.

It should be noted that removing of outliers without recovering or replacing them, will result in a new set of additional missing data, which are different from, e.g., missing measurement data due to failures and faults of measurement systems, or empty recordings. Essentially, these genuine missing data are not outliers, as they do not have a value, which could be assessed in terms of their distance from the reference/regular set of data. Current literature (e.g., [57]) usually classifies missing data in three general types: missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR, also known as nonignorable, NI). The MCAR means the reasons for missing are independent of observable and unobservable parameters and missing occurs entirely at random. The MAR refers that the probability of missing is independent of the missing values and it may be dependent on the observable values, while MNAR or NI occurs when the missingness is caused by the missing values. If missing data are MCAR, then no more new information can be obtained from the missing data and subsequent analysis of the observable data without considering the missing data will not introduce bias. However, MCAR usually does not happen in practice. For MAR and MNAR, the missingness should be taken into consideration, as the data are missing systematically and without them the analysis will be biased [57]. In existing literatures, some of the authors state that their methodologies for filling-in missing data (e.g., multiple imputation) are for MAR only and

not for MNAR. Furthermore, it is suggested that in some cases an application-specific approach will significantly outperform more general models (e.g., multiple imputation). Based on the previous discussion in this paper, the missing data caused by removing outliers due to cut-out effects will result in the nonignorable MNAR-type of missing data (the probability of missing data in the WF fully-operating regime is increasing as the WS approaches the cut-out wind speed, meaning that the values of these data are the reasons why they are missing).

6. Conclusions

This paper presents a comprehensive and systematic review of approaches for the detection and treatment of outliers in processing WT and WF measurements. Starting from the analysis of typically available measurement data, it is first found that most of the previous work was focused on the outliers in WT measurements, while less attention was paid to the analysis of outliers at WF-level, particularly when aggregation of individual WT measurements is used to model the whole WF characteristics. The most widely considered variables are WS and P_{out} . If other recordings are available, the particular attention is given to data normalisation (e.g., power output, wind speed, and turbulence intensity), which is also required by the relevant standard [21]. Afterwards, the paper gives a review of the definitions, characteristics and causes of outliers in WT and WF measurements reported in existing literature, including discussion of outliers due to faults and failures in data acquisition/measurement systems, but also outliers caused by period averaging, cut-out effects, curtailment, etc., introducing simple distinction between true outliers and false outliers. This part of the analysis also includes an examination of data rejection requirements in the relevant standard [21]. Next, the paper reviews methods for the detection of outliers, concluding that the statistical methods (density-based, distance-based, correlation-based, or image-based) are the most popular, while the physical-constraint based methods and combined statistical and physical methods are reported in a relatively small number of references. Additionally, the paper also reviews work aimed at building outlier-insensitive robust statistical models for reducing the effects of outliers without identifying them. Finally, the paper reviews various approaches for assessing the accuracy/success of outlier detection methods, including visual inspection, data elimination rate, comparison of statistical properties, evaluation of PC fitting errors, and data concentration indices. In terms of outlier treatment procedures, it was found that in most of the previous work outliers are simply removed and that only a small number of references treat the outliers as missing data and perform data imputation. This effectively results in additional sets of missing data, where of particular concern are missing data due to elimination of outliers caused by cut-out effects, representing less-probable, but still plausible and intended operating points of a WT/WF.

The main difficulty reported in the literature on outlier detection and treatment is a general lack of information on the actual causes of outliers, e.g., operational logs and similar documentations, which effectively prevent successful, reliable and convincing testing of the proposed outlier detection methods. In different research, the dimensions and resolutions of available measurements may differ. The larger number of data could increase the confidence of the detection of some outliers, but that may also introduce other additional or new outliers.

As discussed, the sources of outliers are many and various, of which some may result in outliers with similar, or even the same patterns. Additionally, some of these outliers may be masked and hidden by the normal data. Therefore, it is sometimes almost impossible to determine the exact causes of outliers, even if their presence may be successfully detected. WT-level outliers and WF-level outliers are different, since some outliers typically occur at a WT-level and not at a WF-level (and vice versa), as the WT and WF operating regimes are generally different. For example, data recorded during the outage of an individual WT are true outliers for the analysis at WT-level, but these data generally should not be excluded from the analysis of operational performance of a WF, i.e., these data are false outliers at WF-level. This review paper did not find standard or unified approaches for treatments of outliers, as different approaches are highly dependent on their purpose, target analysis, or intended use

of the processed measurement data. Generally, this diversity may make different approaches proposed for outlier treatment incomparable. Finally, the occurrences and patterns of some outliers may strongly depend on the specific types of WTs, their locations and variations of their operating/environmental conditions. These factors present additional difficulties for both the provision of the general test benchmark values with labelled data and for the specification of commonly accepted methodologies for the confident identification and adequate processing of outliers.

In terms of the further research, it can be generally concluded that benchmark data sets allowing for a fair assessment and comparison of outlier detection and treatment methods are highly desired. These benchmark data should consider and acknowledge number of factors influencing diversity in various research applications and scenarios, including: (1) available measurement parameters and dimensions; (2) available measurement resolutions; (3) types of WTs; (4) locations of WTs; and (5) operating/environmental conditions of WTs. Furthermore, these benchmark data sets should have labelled data separately for WT-level studies, directly measured WF-level studies and aggregated WF-level studies. Finally, the benchmark data sets could also provide much needed suggestions for further processing and treatment (removal or replacement) of the outliers, depending on the purpose of the research.

Author Contributions: Conceptualization, methodology, validation, analysis, writing and editing, M.Z. and S.Z.D.; software and visualization, M.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by an internal grant from School of Engineering, the University of Edinburgh, grant name: School of Engineering PhD Scholarship.

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

Abbreviations

ANFIS	Adaptive neuro-fuzzy inference
DBSCAN	Density-based spatial clustering of applications with noise
$f(X Y)$	The deterministic relationship between predictor X and response Y
FCM	Fuzzy c-means
FN	False negative
FP	False positive
GARCH	Generalised autoregressive conditional heteroskedasticity
GMCM	Gaussian mixture copula model
GW	Gigawatt
HMM	Hidden Markov model
IQR	Interquartile range
k -CV	k -cross validation
KLOF	Kernel-based local outlier factor
KNN	k -nearest neighbour
kW	Kilowatt
LMedS	Least median of squares
LOF	Local outlier factor
LS-SVR	Least squares support vector regression
MAD	Median absolute deviation
MAR	Missing at random
MCAR	Missing completely at random
MET	Meteorological
Mfr-PC	Manufacturer's power curve
MLP	Multilayer perceptron
MMO	Mathematical morphology operation
MNAR	Missing not at random
NI	Nonignorable
OSTAR	Outlier smooth transition autoregressive

PC	Power curve
PCA	Principal component analysis
P_{out}	Power output
P_{out_max}	Maximum power output
$Prob(X, Y)$	The joint probability of X and Y
$Prob(X Y)$	The conditional probability of X for given Y
SDAE	Stacked denoising autoencoder
STW	Sliding time window
TN	True negative
TP	True positive
WD	Wind direction
WF	Wind farm
WS	Wind speed
WT	Wind turbine

References

1. Papaefthymiou, G.; Kurowicka, D. Using Copulas for Modeling Stochastic Dependence in Power System Uncertainty Analysis. *IEEE Trans. Power Syst.* **2009**, *24*, 40–49. [\[CrossRef\]](#)
2. Fang, D.; Zou, M.; Djokic, S. Probabilistic OPF Incorporating Uncertainties in Wind Power Outputs and Line Thermal Ratings. In Proceedings of the International Conference on Probabilistic Methods Applied to Power Systems (PMAPS), Boise, ID, USA, 24–28 June 2018.
3. Zou, M.; Fang, D.; Djokic, S.; Di Giorgio, V.; Langella, R.; Testa, A. Evaluation of Wind Turbine Power Outputs with and without Uncertainties in Input Wind Speed and Wind Direction Data. *IET Renew. Power Gener.* **2020**. [\[CrossRef\]](#)
4. Watson, S.J.; Xiang, B.J.; Yang, W.; Tavner, P.J.; Crabtree, C.J. Condition Monitoring of the Power Output of Wind Turbine Generators Using Wavelets. *IEEE Trans. Energy Convers.* **2010**, *25*, 715–721. [\[CrossRef\]](#)
5. Zhang, Z.; Sun, Y.; Gao, D.W.; Lin, J.; Cheng, L. A Versatile Probability Distribution Model for Wind Power Forecast Errors and Its Application in Economic Dispatch. *IEEE Trans. Power Syst.* **2013**, *28*, 3114–3125. [\[CrossRef\]](#)
6. Ghofrani, M.; Arabali, A.; Etezadi-Amoli, M.; Fadali, M.S. Energy Storage Application for Performance Enhancement of Wind Integration. *IEEE Trans. Power Syst.* **2013**, *28*, 4803–4811. [\[CrossRef\]](#)
7. Schlechtingen, M.; Santos, I.F.; Achiche, S. Using Data-Mining Approaches for Wind Turbine Power Curve Monitoring: A Comparative Study. *IEEE Trans. Sustain. Energy* **2013**, *4*, 671–679. [\[CrossRef\]](#)
8. Lydia, M.; Kumar, S.S.; Selvakumar, A.I.; Prem Kumar, G.E. A comprehensive review on wind turbine power curve modeling techniques. *Renew. Sustain. Energy Rev.* **2014**, *30*, 452–460. [\[CrossRef\]](#)
9. Carrillo, C.; Obando, M.A.F.; Cidrás, J.; Díaz-Dorado, E. Review of power curve modelling for wind turbines. *Renew. Sustain. Energy Rev.* **2013**, *21*, 572–581. [\[CrossRef\]](#)
10. Hayes, B.P.; Ilie, I.; Porpodas, A.; Djokic, S.Z.; Chicco, G. Equivalent power curve model of a wind farm based on field measurement data. In Proceedings of the 2011 IEEE Trondheim PowerTech, Trondheim, Norway, 19–23 June 2011.
11. Ye, L.; Zhao, Y.; Zeng, C.; Zhang, C. Short-term wind power prediction based on spatial model. *Renew. Energy* **2017**, *101*, 1067–1074. [\[CrossRef\]](#)
12. Xu, M.; Pinson, P.; Lu, Z.; Qiao, Y.; Min, Y. Adaptive robust polynomial regression for power curve modeling with application to wind power forecasting. *Wind Energy* **2016**, *19*, 2321–2336. [\[CrossRef\]](#)
13. Gill, S.; Stephen, B.; Galloway, S. Wind Turbine Condition Assessment Through Power Curve Copula Modeling. *IEEE Trans. Sustain. Energy* **2012**, *3*, 94–101. [\[CrossRef\]](#)
14. Kusiak, A.; Zheng, H.; Song, Z. Models for monitoring wind farm power. *Renew. Energy* **2009**, *34*, 583–590. [\[CrossRef\]](#)
15. Kusiak, A.; Verma, A. Monitoring Wind Farms With Performance Curves. *IEEE Trans. Sustain. Energy* **2013**, *4*, 192–199. [\[CrossRef\]](#)
16. Sarkar, S.; Ajjarapu, V. MW Resource Assessment Model for a Hybrid Energy Conversion System With Wind and Solar Resources. *IEEE Trans. Sustain. Energy* **2011**, *2*, 383–391. [\[CrossRef\]](#)

17. Dhungana, D.; Karki, R. Data Constrained Adequacy Assessment for Wind Resource Planning. *IEEE Trans. Sustain. Energy* **2015**, *6*, 219–227. [\[CrossRef\]](#)
18. Taslimi-Renani, E.; Modiri-Delshad, M.; Elias, M.F.M.; Rahim, N.A. Development of an enhanced parametric model for wind turbine power curve. *Appl. Energy* **2016**, *177*, 544–552. [\[CrossRef\]](#)
19. Wang, S.; Zhang, X.; Ge, L.; Wu, L. 2-D Wind Speed Statistical Model for Reliability Assessment of Microgrid. *IEEE Trans. Sustain. Energy* **2016**, *7*, 1159–1169. [\[CrossRef\]](#)
20. Zhao, Y.; Ye, L.; Wang, W.; Sun, H.; Ju, Y.; Tang, Y. Data-Driven Correction Approach to Refine Power Curve of Wind Farm Under Wind Curtailment. *IEEE Trans. Sustain. Energy* **2018**, *9*, 95–105. [\[CrossRef\]](#)
21. BS EN 61400-12-1:2017. *Wind Energy Generation Systems. Part 12-1: POWER Performance Measurements of Electricity Producing Wind Turbines*; BSI Standards Publication: London, UK, 2017.
22. Kim, H.; Singh, C.; Sprintson, A. Simulation and Estimation of Reliability in a Wind Farm Considering the Wake Effect. *IEEE Trans. Sustain. Energy* **2012**, *3*, 274–282. [\[CrossRef\]](#)
23. Sainz, E.; Llombart, A.; Guerrero, J.J. Robust filtering for the characterization of wind turbines: Improving its operation and maintenance. *Energy Convers. Manag.* **2009**, *50*, 2136–2147. [\[CrossRef\]](#)
24. Villanueva, D.; Feijóo, A. Normal-Based Model for True Power Curves of Wind Turbines. *IEEE Trans. Sustain. Energy* **2016**, *7*, 1005–1011. [\[CrossRef\]](#)
25. Wang, S.; Huang, Y.; Li, L.; Liu, C. Wind turbines abnormality detection through analysis of wind farm power curves. *Measurement* **2016**, *93*, 178–188. [\[CrossRef\]](#)
26. Javadi, M.; Malyscheff, A.M.; Wu, D.; Kang, C.; Jiang, J.N. An algorithm for practical power curve estimation of wind turbines. *CSEE J. Power Energy Syst.* **2018**, *4*, 93–102. [\[CrossRef\]](#)
27. 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. [\[CrossRef\]](#)
28. Stephen, B.; Galloway, S.J.; McMillan, D.; Hill, D.C.; Infield, D.G. A Copula Model of Wind Turbine Performance. *IEEE Trans. Power Syst.* **2011**, *26*, 965–966. [\[CrossRef\]](#)
29. Shen, X.; Fu, X.; Zhou, C. A Combined Algorithm for Cleaning Abnormal Data of Wind Turbine Power Curve Based on Change Point Grouping Algorithm and Quartile Algorithm. *IEEE Trans. Sustain. Energy* **2019**, *10*, 46–54. [\[CrossRef\]](#)
30. Zhao, Y.; Liu, Y.; Wang, R. Fuzzy scalar quantisation based on hidden Markov model and application in fault diagnosis of wind turbine. *J. Eng.* **2017**, *2017*, 2685–2689. [\[CrossRef\]](#)
31. Park, J.; Lee, J.; Oh, K.; Lee, J. Development of a Novel Power Curve Monitoring Method for Wind Turbines and Its Field Tests. *IEEE Trans. Energy Convers.* **2014**, *29*, 119–128. [\[CrossRef\]](#)
32. Yampikulsakul, N.; Byon, E.; Huang, S.; Sheng, S.; You, M. Condition Monitoring of Wind Power System With Nonparametric Regression Analysis. *IEEE Trans. Energy Convers.* **2014**, *29*, 288–299.
33. Xu, X.; Lei, Y.; Li, Z. An Incorrect Data Detection Method for Big Data Cleaning of Machinery Condition Monitoring. *IEEE Trans. Ind. Electron.* **2020**, *67*, 2326–2336. [\[CrossRef\]](#)
34. Cambron, P.; Lepvrier, R.; Masson, C.; Tahan, A.; Pelletier, F. Power curve monitoring using weighted moving average control charts. *Renew. Energy* **2016**, *94*, 126–135. [\[CrossRef\]](#)
35. Kusiak, A.; Zheng, H.; Song, Z. On-line monitoring of power curves. *Renew. Energy* **2009**, *34*, 1487–1493. [\[CrossRef\]](#)
36. Pinson, P.; Madsen, H. Adaptive modelling and forecasting of offshore wind power fluctuations with Markov-switching autoregressive models. *J. Forecast.* **2012**, *31*, 281–313. [\[CrossRef\]](#)
37. Mangalova, E.; Agafonov, E. Wind power forecasting using the k-nearest neighbors algorithm. *Int. J. Forecast.* **2014**, *30*, 402–406. [\[CrossRef\]](#)
38. Lowery, C.; Malley, M.O. Impact of Wind Forecast Error Statistics Upon Unit Commitment. *IEEE Trans. Sustain. Energy* **2012**, *3*, 760–768. [\[CrossRef\]](#)
39. Liu, Y.; Liu, Q.; Han, S.; Yan, J.; Li, L. The detection method of wind turbine operation outliers based on multi-dimensional cluster. In Proceedings of the 8th Renewable Power Generation Conference (RPG 2019), Shanghai, China, 24–25 October 2019.
40. Zheng, L.; Hu, W.; Min, Y. Raw Wind Data Preprocessing: A Data-Mining Approach. *IEEE Trans. Sustain. Energy* **2015**, *6*, 11–19. [\[CrossRef\]](#)
41. Wang, Y.; Infield, D.G.; Stephen, B.; Galloway, S.J. Copula based model for wind turbine power curve outlier rejection. *Wind Energy* **2014**, *17*, 1677–1688. [\[CrossRef\]](#)

42. Hu, Y.; Qiao, Y.; Liu, J.; Zhu, H. Adaptive Confidence Boundary Modeling of Wind Turbine Power Curve Using SCADA Data and Its Application. *IEEE Trans. Sustain. Energy* **2019**, *10*, 1330–1341. [\[CrossRef\]](#)
43. Ouyang, T.; Kusiak, A.; He, Y. Modeling wind-turbine power curve: A data partitioning and mining approach. *Renew. Energy* **2017**, *102*, 1–8. [\[CrossRef\]](#)
44. Mohammadpourfard, M.; Sami, A.; Weng, Y. Identification of False Data Injection Attacks With Considering the Impact of Wind Generation and Topology Reconfigurations. *IEEE Trans. Sustain. Energy* **2018**, *9*, 1349–1364. [\[CrossRef\]](#)
45. Yesilbudak, M. Partitional clustering-based outlier detection for power curve optimization of wind turbines. In Proceedings of the 2016 IEEE International Conference on Renewable Energy Research and Applications (ICRERA), Birmingham, UK, 20–23 November 2016.
46. Zhou, Y.; Hu, W.; Min, Y.; Zheng, L.; Liu, B.; Yu, R.; Dong, Y. A semi-supervised anomaly detection method for wind farm power data preprocessing. In Proceedings of the 2017 IEEE Power & Energy Society General Meeting, Chicago, IL, USA, 16–20 July 2017.
47. Sun, Z.; Sun, H. Stacked Denoising Autoencoder With Density-Grid Based Clustering Method for Detecting Outlier of Wind Turbine Components. *IEEE Access* **2019**, *7*, 13078–13091. [\[CrossRef\]](#)
48. Ye, X.; Lu, Z.; Qiao, Y.; Min, Y.; Malley, M.O. Identification and Correction of Outliers in Wind Farm Time Series Power Data. *IEEE Trans. Power Syst.* **2016**, *31*, 4197–4205. [\[CrossRef\]](#)
49. Long, H.; Sang, L.; Wu, Z.; Gu, W. Image-Based Abnormal Data Detection and Cleaning Algorithm via Wind Power Curve. *IEEE Trans. Sustain. Energy* **2020**, *11*, 938–946. [\[CrossRef\]](#)
50. Wan, Y.-H.; Ela, E.; Orwig, K. Development of an equivalent wind plant power curve. In Proceedings of the Windpower 2010 Conference & Exhibition, Dallas, TX, USA, 23–26 May 2010.
51. Paiva, L.T.; Veiga, R.C.; Palma, J.M.L.M. Determining wind turbine power curves based on operating conditions. *Wind Energy* **2014**, *17*, 1563–1575. [\[CrossRef\]](#)
52. Chen, H.; Li, F.; Wang, Y. Wind power forecasting based on outlier smooth transition autoregressive GARCH model. *J. Mod. Power Syst. Clean Energy* **2018**, *6*, 532–539. [\[CrossRef\]](#)
53. Morshedizadeh, M.; Kordestani, M.; Carriveau, R.; Ting, D.S.; Saif, M. Power production prediction of wind turbines using a fusion of MLP and ANFIS networks. *IET Renew. Power Gener.* **2018**, *12*, 1025–1033. [\[CrossRef\]](#)
54. Härdle, W.K.; Simar, L. Principal Components Analysis. In *Applied Multivariate Statistical Analysis*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 319–358.
55. Albers, A. Turbulence and Shear Normalisation of Wind Turbine Power Curve. In Proceedings of the European Wind Energy Conference and Exhibition (EWEC), Warsaw, Poland, 20–23 April 2010.
56. Albers, A.; Jakobi, T.; Rohden, R.; Stoltenjohannes, J. Influence of meteorological variables on measured wind turbine power curves. In Proceedings of the European Wind Energy Conference and Exhibition 2007 (EWEC), Milan, Italy, 7–10 May 2007.
57. King, G.; Honaker, J.; Joseph, A.; Scheve, K. Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation. *Am. Political Sci. Rev.* **2001**, *95*, 49–69. [\[CrossRef\]](#)
58. Zou, M.; Fang, D.; Djokic, S.; Hawkins, S. Assessment of wind energy resources and identification of outliers in on-shore and off-shore wind farm measurements. In Proceedings of the 3rd International Conference on Offshore Renewable Energy (CORE), Glasgow, UK, 29–30 August 2018.
59. Gayo, J.B. *Reliability focused research on optimizing Wind Energy Systems Design, Operation and Maintenance: Tools, Proof of Concepts, Guidelines & Methodologies for a New Generation (ReliaWind)*; Gamesa Innovation and Technology: Madrid, Spain, 2011.
60. Vestas. *Vestas Wind Systems A/S, Report 950010.R1: General Specification V90–3.0 MW*; Vestas: Aarhus, Denmark, 2013.
61. Gupta, M.; Gao, J.; Aggarwal, C.C.; Han, J. Outlier Detection for Temporal Data: A Survey. *IEEE Trans. Knowl. Data Eng.* **2014**, *26*, 2250–2267. [\[CrossRef\]](#)
62. Chandola, V.; Banerjee, A.; Kumar, V. Anomaly detection: A survey. *ACM Comput. Surv.* **2009**, *41*, 15. [\[CrossRef\]](#)
63. Tukey, J.W. *Exploratory Data Analysis*; Reading, Mass.; Addison-Wesley Pub. Co.: Boston, MA, USA, 1977.

64. Lange, M. On the Uncertainty of Wind Power Predictions—Analysis of the Forecast Accuracy and Statistical Distribution of Errors. *J. Sol. Energy Eng.* **2005**, *127*, 177–184. [\[CrossRef\]](#)
65. Zhou, Y.; Wan, A.T.K.; Xie, S.; Wang, X. Wavelet analysis of change-points in a non-parametric regression with heteroscedastic variance. *J. Econom.* **2010**, *159*, 183–201. [\[CrossRef\]](#)
66. Ester, M.; Kriegel, H.-P.; Sander, J.; Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. In Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, Portland, OR, USA, 2–4 August 1996; AAAI Press: Portland, Oregon, 1996; pp. 226–231.
67. Gao, J.; Hu, W.; Li, W.; Zhang, Z.; Wu, O. Local Outlier Detection Based on Kernel Regression. In Proceedings of the 2010 20th International Conference on Pattern Recognition, Istanbul, Turkey, 23–26 August 2010.
68. Breunig, M.M.; Kriegel, H.-P.; Ng, R.T.; Sander, J. LOF: Identifying density-based local outliers. In Proceedings of the 2000 ACM SIGMOD international conference on Management of data, Renton, WA, USA, 19–21 October 2000; ACM: Dallas, TX, USA, 2000; pp. 93–104.
69. Ruiyu, J.; Li, Y. K-means algorithm of clustering number and centers self-determination. *Comput. Eng. Appl.* **2018**, *54*, 152–158.
70. Khodayar, M.; Kaynak, O.; Khodayar, M.E. Rough Deep Neural Architecture for Short-Term Wind Speed Forecasting. *IEEE Trans. Ind. Inform.* **2017**, *13*, 2770–2779. [\[CrossRef\]](#)
71. Lu, J.; Zhu, Q. An Effective Algorithm Based on Density Clustering Framework. *IEEE Access* **2017**, *5*, 4991–5000. [\[CrossRef\]](#)
72. Wu, B.; Wilamowski, B.M. A Fast Density and Grid Based Clustering Method for Data With Arbitrary Shapes and Noise. *IEEE Trans. Ind. Inform.* **2017**, *13*, 1620–1628. [\[CrossRef\]](#)
73. Witten, I.H.; Frank, E.; Hall, M.A. Chapter 3—Output: Knowledge Representation. In *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed.; Witten, I.H., Frank, E., Hall, M.A., Eds.; Morgan Kaufmann: Boston, MA, USA, 2011; pp. 61–83.
74. Mitra, A. *Fundamentals of Quality Control. and Improvement*, 3rd ed.; John Wiley & Sons, Inc: Hoboken, NJ, USA, 2008.
75. Sklar, A. *Fonctions de Répartition à n Dimensions et Leurs Marges [N-Dimensional Joint and Marginal Distribution Functions]*; Université Paris 8: Paris, France, 1959; pp. 229–231.
76. Tewari, A.; Giering, M.J.; Raghunathan, A. Parametric Characterization of Multimodal Distributions with Non-gaussian Modes. In Proceedings of the 2011 IEEE 11th International Conference on Data Mining Workshops, Vancouver, BC, Canada, 11 December 2011.
77. Brown, R.G.; Meyer, R.F. The Fundamental Theorem of Exponential Smoothing. *Oper. Res.* **1961**, *9*, 673–685. [\[CrossRef\]](#)
78. Markowski, C.A.; Markowski, E.P. Conditions for the Effectiveness of a Preliminary Test of Variance. *Am. Stat.* **1990**, *44*, 322–326.
79. Bezdek, J.C.; Ehrlich, R.; Full, W. FCM: The fuzzy c-means clustering algorithm. *Comput. Geosci.* **1984**, *10*, 191–203. [\[CrossRef\]](#)
80. Leys, C.; Ley, C.; Klein, O.; Bernard, P.; Licata, L. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *J. Exp. Soc. Psychol.* **2013**, *49*, 764–766. [\[CrossRef\]](#)
81. Haralick, R.M.; Sternberg, S.R.; Zhuang, X. Image Analysis Using Mathematical Morphology. *IEEE Trans. Pattern Anal. Mach. Intell.* **1987**, *9*, 532–550. [\[CrossRef\]](#)
82. Ming-Kuei, H. Visual pattern recognition by moment invariants. *IRE Trans. Inf. Theory* **1962**, *8*, 179–187. [\[CrossRef\]](#)
83. Lloyd, S.P. Least squares quantization in PCM. *IEEE Trans. Inf. Theory* **1982**, *28*, 129–136. [\[CrossRef\]](#)
84. Suykens, J.A.K.; De Brabanter, J.; Lukas, L.; Vandewalle, J. Weighted least squares support vector machines: Robustness and sparse approximation. *Neurocomputing* **2002**, *48*, 85–105. [\[CrossRef\]](#)
85. David, H.A. Early Sample Measures of Variability. *Stat. Sci.* **1998**, *13*, 368–377. [\[CrossRef\]](#)
86. Lin, Q.; Wang, J. Vertically Correlated Echelon Model for the Interpolation of Missing Wind Speed Data. *IEEE Trans. Sustain. Energy* **2014**, *5*, 804–812. [\[CrossRef\]](#)
87. Yang, Z.; Liu, Y.; Li, C. Interpolation of missing wind data based on ANFIS. *Renew. Energy* **2011**, *36*, 993–998. [\[CrossRef\]](#)

88. Zhang, Y.; Kim, S.; Giannakis, G.B. Short-term wind power forecasting using nonnegative sparse coding. In Proceedings of the 2015 49th Annual Conference on Information Sciences and Systems (CISS), Baltimore MD, USA, 18–20 March 2015.
89. Efron, B. Missing Data, Imputation, and the Bootstrap. *J. Am. Stat. Assoc.* **1994**, *89*, 463–475. [[CrossRef](#)]
90. Takagi, H.; Hayashi, I. NN-driven fuzzy reasoning. *Int. J. Approx. Reason.* **1991**, *5*, 191–212. [[CrossRef](#)]
91. Jang, J.R. ANFIS: Adaptive-network-based fuzzy inference system. *IEEE Trans. Syst. Man and Cybern.* **1993**, *23*, 665–685. [[CrossRef](#)]

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).