BUNDESAMT FÜR
SEESCHIFFFAHRT
UND
HYDROGRAPHIE

RAVE Offshoreservice

Quality Control of RAVE Measurements from AV00, AV04, AV05, AV07–AV12 and FINO1

Bundesamt für Seeschifffahrt
und Hydrographie

Version 1.0 of November 2019

Contact

Federal Maritime and Hydrographic Agency
Herr Kai Herklotz
Bernhard-Nocht-Str. 78, 20359 Hamburg, Germany
+49 (0)40 3190-3230
rave-forschungsarchiv@bsh.de

UL International GmbH
Dr. Tom Neumann
Ebertstr. 96, 26382 Wilhelmshaven, Germany
+49 (0)4421 4808-814
T.Neumann@dewi.de

DNV GL (GL Garrad Hassan Deutschland GmbH)
Hans-Peter Link
Sommerdeich 14 b, 25709 Kaiser-Wilhelm-Koog, Germany
+49 (0)4856 901-46
Hans-Peter.Link@dnvgl.com

RAVE
RESEARCH AT ALPHA VENTUS

UL

DNV·GL

Table of Contents

Revision History

Date	Changed Chapters	Revision Description	Notes
26. November 2019		Original Document Published	

1 Introduction

The following describes the first version of the quality-control procedure for the measurements collected as part of the RAVE project[1] at the alpha ventus wind farm in the North Sea by UL International GmbH and DNV GL. These measurements include structural (AV07 and AV08) and machine data (AV07–AV12) for the Adwen M5000 wind turbines, structural and machine data (AV04 and AV05) for the Senvion 5M wind turbines, electrical measurements at the AV00 transformer, and meteorological data from the FINO1 platform west of the AV04 wind turbine, forming part of the new RAVE database at the BSH. Further data within the database are the oceanographic data collected by the BSH and described in a separate report[2]. Examples of the types of sensors considered here are given in Table 1.

The quality-control tests were originally conceived as being independent of the sensor type and wind-turbine status, with the only required user inputs being the specific parameters for each test. This is because the turbine status signal may not always be available, and there may be missing sensors or measurement modules preventing consistent formal side-by-side comparisons. Further evaluation of the physical plausibility of the derived loads based on the structural measurements is included in a separate work package of the RAVE project.

Table 1 – Sensor types operated within the RAVE project, including the sampling frequency.

Type	Example	Frequency
Structural	Strain gauge	50 Hz
Machine	Pitch Angle	50 Hz
Meteorology	Wind Speed	1 Hz, 20 Hz
Electrical	Voltage	0.2* Hz

* Can be increased to the kHz range.

[1] The research project RAVE Offshoreservice is funded by the German Federal Ministry of Economic Affairs and Energy on the basis of a decision by the German Bundestag and coordinated by Fraunhofer IWES.

[2] BSH (2019). Real-Time Data Quality Control: In Situ Surface Waves, Version 1.0, August 2019.

2 Background Quality Control

In the original RAVE database operated by OFFIS e. V., only two of the quality tests described here were employed: the range and visual evaluation tests (see Table 3). An extension of the quality-control strategy was envisioned at the start of the research project RAVE Offshoreservice (Phase 3), since the philosophy of the RAVE database is to include as much data as possible, even of potentially poor quality, with the hope that some information may be prove useful for future users. Therefore, with the transfer of the old database to and implementation of the new database at the BSH, it was decided to take the opportunity to add the additional tests described here.

3 Brief Review

Two of the key sources serving as a basis for the quality-control procedure include the IEC Standard 61400-13[3] and the guide for high-frequency micrometeorological data by Vickers and Mahrt (1997)[4]. The former source lists the recommended types of time-series properties to verify while the latter gives more details as to how to identify specific problems with time series. Section 9.2.1 of IEC 61400-13 recommends checking for the violation of sensor opera-tional limits, sensor drift, false calibration, expected eigenfrequencies, noise, spikes, flat lines, and significant differences between comparable sensors. Those checks that can be easily automated are included in this procedure, while others closer related to the physical properties are either included in a "manual test" or in a separate project work package. The detection of spikes and other issues in raw time series is the focus of Vickers and Mahrt (1997), who, while focusing on micrometeorological flux sampling, still provide a useful basis for the detection of similar issues arising in structural measurements (for example strain gauges), not to mention the meteorological measurements also handled in the RAVE database.

[3] Wind Turbines – Part 13: Measurement of Mechanical Loads

[4] Vickers, D., and Mahrt, L. "Quality Control and Flux Sampling Problems for Tower and Aircraft Data." Journal of Atmospheric and Oceanic Technology 14 (1997): 512–526.

4 Metadata

Each 10-min time series of the RAVE database is associated with the metadata listed in Table 2 from which much of the data quality may be inferred. The Flag and Detailed Flag give the explicit results of the quality-control procedure described in the following sections. The Percentage gives the recorded number of data points, and may even exceed 100 % if the actual sampling rate slightly exceeds the nominal sampling rate (for example 50.001 Hz). Basic statistical information calculated from the full time series includes the mean value (Mean), the standard deviation (Stddev), and the minimum (Min) and maximum (Max) values within a 10-min time series. The following is an incomplete list of data qualities that *may* be inferred from the metadata:

- Sensor failure: The values of Min/Max for strain or acceleration are equal.
- Measurement-system or sensor dropout: Percentage << 100 %.
- Spike(s) or signal dropout: Mean value differs significantly from the Min/Max values.
- Noisy data: Value of Stddev tends to the absolute value of the range (= |Max – Min|).

Table 2 – The metadata associated with each 10-min time series in the RAVE database

Metadata	Description
Flag	Value (1/0) indicating whether at least one threshold of the quality-control procedure is violated or if procedure not performed (9).
Detailed Flag	Values (1/0) indicating whether the threshold of each test of the quality-control procedure was exceeded (see Table 3) or if test not performed (9).
Percentage	Fraction of actual data as a percentage of a complete time series
[Mean, Stddev, Min, Max]	Basic statistical information (mean, standard deviation, minimum and maximum measured values)

5 Quality Test Summary

Table 3 summarizes the individual tests carried out on each of the 10-min time series collected within the RAVE project, including the testing order 1–7, the threshold parameters employed in each test, and a basic description of each test. The positions 8–16 are reserved for further tests to be included during the course of the RAVE project. The results of each test are delivered as a detailed flag consisting of zeros (if threshold not violated), ones (threshold violated) or nines (test not conducted) for each position, and summarized as a master flag of a single zero (no threshold violated), one (at least one threshold violated) or nine (no test performed).

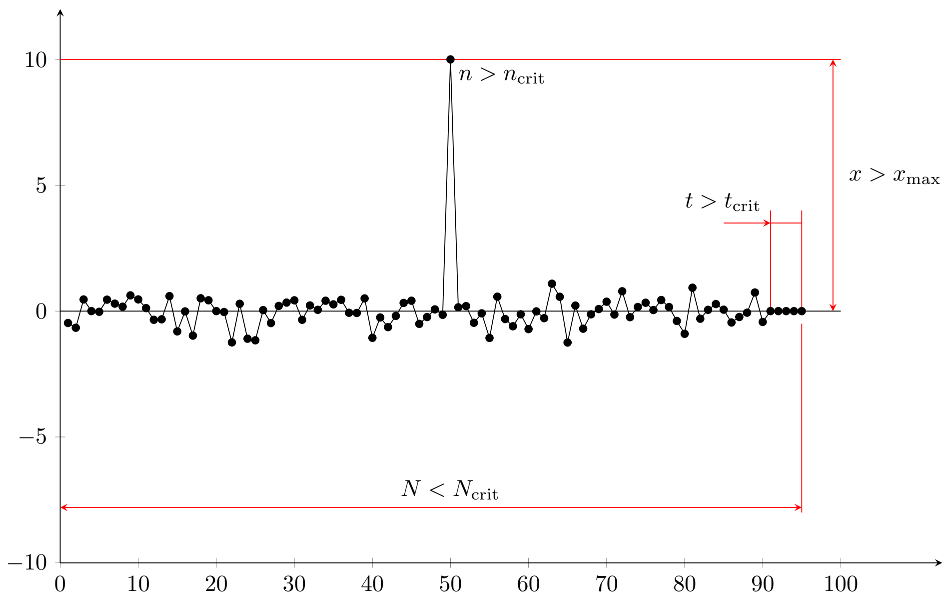

Figure 1 – Artificial time series of length $N <$ the critical signal length N_{crit} illustrating the thresholds employed in the quality-control procedure of the RAVE data (see also Table 3). Other thresholds are the partial flat line length t_{crit}, the number of spikes n_{crit}, and the measurement range x_{max}, where x_{max} is the maximum allowable value.

An illustration of the thresholds of an arbitrary artificial signal is provided in Figure 1. Here, ideally $N = 100$ samples would have been collected were it not for a signal dropout beginning at the sample $i = 90$, with data taking a constant value $x = 0$, followed by a premature end to the time series. An outlier in the time series is also present at the sample position $i = 50$, which also exceeds the measurement range.

Table 3 – Summary of the individual tests comprising the RAVE quality-control procedure. Values of the thresholds are given in Appendix A.

Position	Test Type	Meaning	Thresholds	Description
1	Length	Reduced data length	$N_{crit} > \| N \pm L_{tol} \|$	Time series deviating from the expected length N_{crit} by some tolerance (e. g. L_{tol} = 3 points)
2	Flat Line	Constant Signal	$\sum (x_i - x_{i+1} = 0) = N - 1$	All values x_i in the time series of length N are the same.
3	Flat Line	Partially Constant Signal	t_{crit}	Constant values for a period of > t_{crit} seconds (e. g. signal dropouts)
4	Pre-defined Limits	Signal Range	$[x_{min}\ x_{max}]$	At least one value outside the given range (e. g. ±10 V, –30°C–100°C)
5	Spike	Spike events exceeded	n_{crit}	Number of spikes n found in signal exceeds critical value n_{crit}.
6	Spike	Low Correlation	r_{crit}	Despiked signal poorly correlated with uncorrected signal based on the correlation coefficient r.
7	Visual	–	–	Manual identification of signal problems
8–16	–	– Spare –	–	Further tests included here.

Assuming the following thresholds: number of spikes $n_{crit} = 0$, measurement range $[x_{max}\ x_{min}] = \pm 10$, critical number of consecutive values $t_{crit} = 3$ samples, and a critical data length $N_{crit} = 99$ samples, then the signal would receive the following detailed flag assuming the test order displayed in Table 3:

F_d = [1, 0, 1, 1, 1, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9],

and a master flag

$F_M = 1$, since $\max(F_d) = 1$.

The positions 6 and 7 marked with values of “9” (red), corresponding to the spike and “visual” test, respectively, are explained in detail below. The further values of 9 are spare positions.

It should be emphasized that a flag of “1” or “0” does not necessarily imply “bad” or “good” data, respectively, but rather that an event has been detected. The sensor type and the operational status of the wind turbine need to be considered together with the value of the flag and the type of test to which it belongs. For example, an accelerometer with the detailed flag F_d = [0, 1, 0, 0, 0, 0, 0, …] (flat line detected) is probably bad, but the detailed flag F_d = [0, 0, 1, 0, 0, 0, 0, …] (partial flat line detected) for the turbine status signal when the turbine is fully operational may indicate problems with the signal since one expects a full flat line in such a case. There is an extra mechanism in the quality-control procedure to highlight truly bad data in position 7 which is to be manually marked by the measurement institute. For example, the detailed flag F_d = [0, 0, 0, 0, 0, 0, 1, …] for a signal with position 7 explicitly marked "1" would indicate some malfunction that has also been recognized by the measurement institute.

6 Description of the Quality Tests

A description of each individual test is provided here. The testing sequence and the overall strategies are the same for both UL and DNV GL (see Table 1), with some differences in algorithms depending on the sensor. The specific values of the thresholds used by each institute are presented in Appendix A. The formal tests described first are testing for essentially a single critical parameter within a time series. Spike testing described next requires the signal to be processed first with a despiking algorithm after which formal tests are applied (for example spike counting). A brief description of the visual and future planned tests for the spare positions is finally given.

6.1 Formal Tests

6.1.1 Flag Position 1: Length (N_{crit})

Background: As the sampling rate of structural data within the RAVE project is 50 Hz, each 10-min times series should have an expected data length N = 30,000 samples. Exceptions to this are the meteorological data (wind speed, temperature) collected at the FINO1 platform (1 Hz, 20 Hz), the electrical data (voltage, power) at the transformer, and the internal temperature and relative humidity sensors within the AV04 and AV05 wind turbines. Time series of length less than this value N = 30,000 may arise because of park outages or system restarts; time series of length more or less than this value arise because of inaccuracies associated with the hardware of the data logger. In the original OFFIS RAVE database, datasets of length significantly deviating from 100 % were not included. As the new BSH database may contain 10-min time series of any length deviating from the value of N, a test is employed to indicate deviations from 100 % within a tolerance of the normal variance of the hardware. A value of zero for the quality flag but a data percentage deviating from 100 % in the metadata for a 10-min time series indicates there was some fluctuation in the data length but within the defined tolerance. The actual value of N is available as a percentage in the metadata (see Table 2).

Table 4 – Data Length Test

Code	Definition
0	$N \leq \lvert 100\% \pm L_{tol} \rvert = N_{crit}$
1	$N > \lvert 100\% \pm L_{tol} \rvert = N_{crit}$
9	Test not conducted

6.1.2 Flag Position 2: Flat Line

Background: Depending on the sensor, a constant value throughout the time series may signify a defective sensor. Exceptions are some turbine control or machine data if the turbine status remains constant throughout the 10-min time series, or electrical measurements if the wind farm/turbine should be disconnected from the electrical grid for some reason. If the wind turbine is operational throughout a 10-min time series, but the turbine status signal should vary, and hence a value of "0" results in this case, this may also indicate sensor malfunction. Structural measurements should, even if the wind speed is zero, give some detectable deviation in values within a time series.

Procedure: The differenced time series is calculated $x_i - x_{i+1}$ and equated with zero; if all values $x_i - x_{i+1}$ are the same, their summation equates with the signal length minus one.

Table 5 – Flat Line Test

Code	Definition
0	$(\sum(x_i - x_{i+1}) == 0) \neq N - 1$
1	$(\sum(x_i - x_{i+1}) == 0) = N - 1$
9	Test not conducted

6.1.3 Flag Position 3: Partial Flat Line (t_{crit})

Background: Partial flat lines or signal dropouts refer to periods within a time series having the same value, and can be due various reasons, such as communication interruptions between the sensor and the data-logger system, an excessive sampling rate, insufficient bit resolution, or even sensor malfunction. For structural signals such as strain gauges, small changes should still be present even at low wind speeds, but meteorological or control signals may have multiple partial flat lines as these signals are not subject to rapid changes within a 10-min time period.

Procedure: Calculate a moving standard deviation of window width t_{crit} through the differenced time series $x_i - x_{i+1}$. Any zero values in the resulting time series indicates a run of length $\geq t_{crit}$ of the same values.

Table 6 – Partial Flat Line Test

Code	Definition
0	Maximum run length $< t_{crit}$
1	Maximum run length $\geq t_{crit}$
9	Test not conducted

6.1.4 Flag Position 4: Measurement Range

Background: Sample values x_i must be constrained within the measurement or physical range dictated by the sensor type or physical quantity determined, respectively. Values outside the signal measurement range (e. g. ±10 V) indicate either a defective or problematic sensor. More stringent thresholds may be set for sensors with quantities of known physical limits, such as the wind speed (for example, > 40 m s^{-1}) or temperature (for example, > 50 °C), but such physical thresholds are less applicable for structural data, such as strain gauges and accelerometers, where significant deviations from the expected loads or accelerations may even indicate component damage.

Procedure: Data are flagged if the minimum or maximum values in a time series are found outside the defined ranges.

Table 7 – Measurement Range Test

Code	Definition
0	$\min(x_i) > x_{\min}$ and $\max(x_i) < x_{\max}$
1	$\min(x_i) \le x_{\min}$ or $\max(x_i) \ge x_{\max}$
9	Test not conducted

6.2 Testing for Spikes and Signal Noise

Background: Outliers within a signal may indicate a malfunctioning sensor or the presence of non-physical disturbances of electrical or electronic origin potentially impacting the statistical reliability of the time series. Two criteria (spike counting and correlation) are employed to identify the nature of the spike phenomena detected by the despiking algorithm whose procedure is first described.

Procedure: Signals are tested for the presence of outliers by applying a moving mean of window width w through the differenced $x_i - x_{i+1}$ time series, with the absolute differenced values exceeding some multiple m of the standard deviation σ within the window removed and replaced by linear interpolation in the signal x_i. The differenced signal $x_i - x_{i+1}$ is evaluated for outliers rather than the signal x_i in order to identify spikes embedded within sinusoidal signals or to remove physical non-stationarities that may otherwise have been identified as spikes. For example, non-stationary turbine behaviour may yield particularly spike-like behaviour during start-up and shutdown operations of the wind turbine, whereas the differenced signal is more consistent with either operational condition of the wind turbine (by essentially focusing on the higher frequency part of the signal).

In particularly noisy signals having a large value of the standard deviation σ in the differenced signal relative to the mean value μ, the spike-detection algorithm fails to detect any outliers although the data may be of poor quality. More aggressive despiking procedures to handle particularly noisy data are currently being tested for inclusion in futures versions of the quality-control procedure.

The resulting despiked signal is then evaluated using the following two methods:

1. counting the number of spikes (Position 5), and
2. comparing the original signal with the despiked signal through the Pearson correlation coefficient r (Position 6).

6.2.1 Flag Position 5: Spike Counting

Background: The intention of the spike-counting procedure is to distinguish isolated outliers (archetypal spikes) from clusters of outliers (for example drop-outs or other noisy behaviour).

Procedure: A couple of different strategies are deployed for the identification of archetypal spikes in the despiked signal:

i. A spike is the midpoint of the values of the gradient $x_{i-1} - x_{i+1}$ (red curve in Figure 2) of the signal x_i (black points) lying > $m\underline{\sigma}$ (solid red lines), which is some factor m times the standard deviation $\underline{\sigma}$ of the differenced despiked signal $\underline{x}_i - \underline{x}_{i+1}$ (an underline denotes a despiked signal).

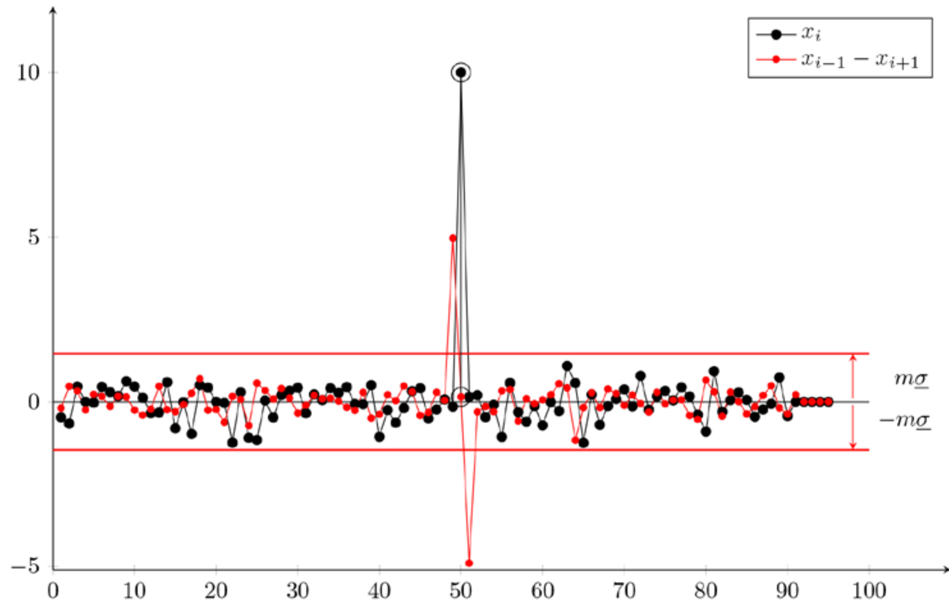

Figure 2 – Artificial signal x_i with a single archetypal spike (black) and the gradient $x_{i-1} - x_{i+1}$ of the corrected signal (red). The midpoint of the points > $m\underline{\sigma}$ gives the spike (circles) at x_{50}.

ii. Find all points in the differenced signal $x_i - x_{i+1}$ outside the range > $m\underline{\sigma}$ and calculate their sign (+/− 1) (Figure 3 top, red curve). Consider only isolated points with a change in sign representing the upstroke/downstroke (or vice versa) of the spike. The archetypal spike in the signal x_i has values before and after < 10 % the magnitude of the spike (Figure 3 bottom), corresponding to the second circled point in Figure 3 (top).

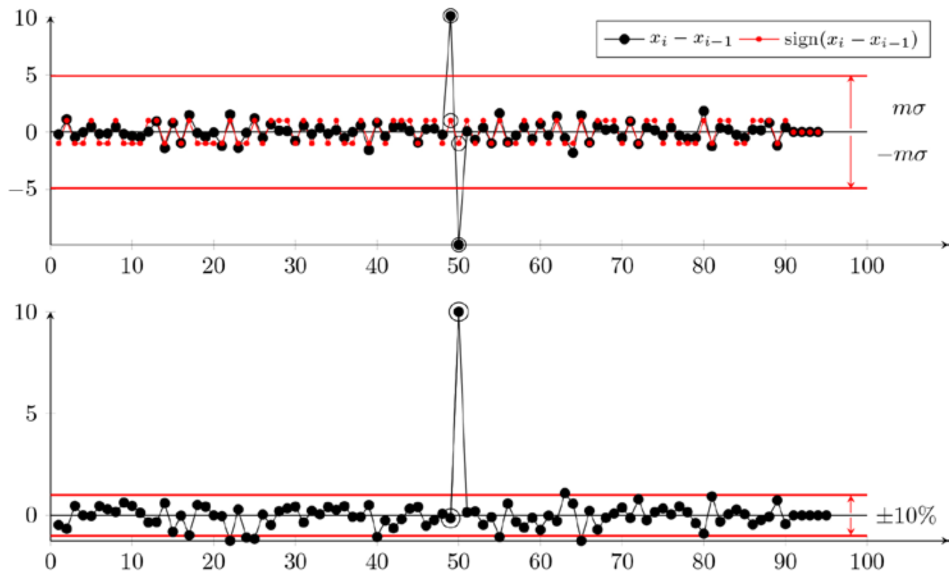

Figure 3 – Top: artificial differenced signal $x_i - x_{i+1}$ with the identified outliers lying > $m\underline{\sigma}$ the mean value within a window and the sign (+/– 1) of this signal (red). Bottom: artificial signal x_i and the range (red) of magnitude ±10 % the value of the spike.

The total number of spikes n are counted within a 10-min time series, which is then flagged if the spike count n exceeds a critical value n_{crit}.

Table 8 – Spike Counting Test

Code	Definition
0	$n < n_{crit}$
1	$n \geq n_{crit}$
9	Test not conducted

6.2.2 Flag Position 6: Correlation

Background: Since spike-like behaviour consisting of clusters of outliers may arise (e. g. Figure 4) different from the strict definition of an archetypal, isolated spike as assessed in Test 5, a separate test is deployed to detect such events. Moreover, the strict definition of a spike described above would fail to detect certain noisy phenomena (for example consecutive spikes).

Procedure: Calculate the correlation coefficient

$$r = N^{-1} \sum(x_i - \mu_x)(\underline{x}_i - \underline{\mu}_x) / (\sigma_x \underline{\sigma}_x)$$

between the raw signal x_i and the despiked signal $\underline{x}_i$. Here, an underline (_) denotes the despiked signal, with μ denoting a mean value, and σ a standard deviation. The value of r may theoretically range from –1 indicating perfect negative correlation, to zero indicating no correlation, and one indicating perfect correlation. For this test, $0 \le r \le 1$ (no negative correlations) as any detected outliers are replaced by the linear interpolation of existing values, making a negative correlation impossible; if no outliers are found, then $r = 1$. Time series are flagged if $r < r_{crit}$, where r_{crit} is the defined threshold. Other statistical parameters, such as the root-mean-squared difference or the bias would also potentially serve the same purpose of assessing spiked data, and may be pursued further in later versions of the quality-control procedure.

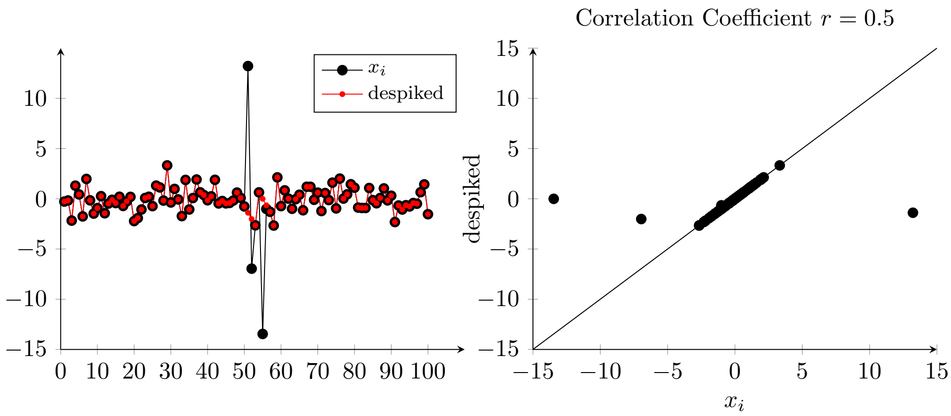

Figure 4 – Left: Time series of the signal x_i (black) affected by a cluster of spikes which are removed by "despiking" (red). Right: scatter plot of the despiked versus raw signals and the value of the correlation coefficient for these two signals.

Table 9 – Spike Correlation Test

Code	Definition
0	$r > r_{crit}$
1	$r > r_{crit}$
9	Test not conducted

6.3 Extra Tests

6.3.1 Flag Position 7: Visual Evaluation

Many signal problems are sensor and position dependent, making it difficult to formalize general tests for the hundreds of sensors in the RAVE database based on a few simple rules. One example is the lack of appearance of peaks at the expected excitation or natural frequencies in the signal spectra, which may indicate a sensor malfunction or a change in structural behaviour. However, as the appearance of these frequency peaks is also dependent on the turbine status, formal tests would involve consideration for the individual sensors and the status of the wind turbine together, which we wish to avoid, since the turbine status signal is not always available, and the status signal itself may suffer from poor quality. If the value of this position is "1", contact the measurement institute for further details regarding the identified problems.

Table 10 – Visual Evaluation Test

Code	Definition
0	Visual evaluation performed and no errors found
1	Visual or manual evaluation reveals problems detected in times series
9	Test not conducted

7 Summary

The first version of the quality-control procedure for the RAVE data has been described, and applies to measurements collected at the AV04/AV05 (Senvion, DNV GL) and AV07–AV12 (Adwen, UL) wind turbines (50 Hz data), the FINO1 (UL) meteorological platform (1 Hz, 20 Hz), and the electrical measurements from the transformer AV00 (UL, 0.2 Hz). Therefore, the intention was to make the procedure as general as possible (not sensor specific). Seven tests have been proposed, with the number of test to be extended to a maximum of 16 tests within the new RAVE database at the BSH in further revisions.

The formal tests are first carried out to highlight specific issues with the time series, including its length (Test 1), whether any values within the time series are different (Test 2), if a partial flat line is present (Test 3), and if the defined measurement range has been exceeded (Test 4). A despiking algorithm is then applied to each time series for the detection of archetypal spikes (Test 5) and spike clusters (Test 6). The visualization test (Test 7) is included to detect problems difficult to formalize.

The results of each test are delivered as a detailed flag consisting of values of zeros (if threshold not violated), ones (threshold violated) or nines (test not conducted) for each position, and summarized as a master flag of a single zero (no threshold violated), one (at least one threshold violated), or nine (no test conducted). Since it was thought difficult to generalize a few simple rules for hundreds of different types of sensors, a flag of “1” or “0” here does not necessarily imply “bad” or “good” data, respectively, but simply that an event has been detected. The sensor type and the operational status of the wind turbine need to be considered together with the value of the flag and the type of test to which the flag belongs.

Further useful tests would be the explicit identification of bad resolution (quantization), a test for the identification of noisy behaviour (whether due to spikes or randomness), more sophisticated algorithms for correcting spike events of multiple points, the identification of signal drift in time or a sudden change in the value of the offset. Machine-learning techniques will also be explored in the next phase of the RAVE project.

Appendix A – Values of Test Parameters

The values of the test parameters defined in Table 3 are given in Table 11. The data have been grouped according to the sampling frequency, data location and/or type.

Table 11 – Values of the thresholds and test parameters defined in Table 3 for AV00, AV04/AV05, AV07–AV12 and FINO1 data at alpha ventus (see Section 6.2).

Parameter (sampling, frequency, Institute)	AV04/AV05 (50 Hz, DNV GL)	AV07 (50 Hz, UL)	AV08 (50 Hz, UL)	AV07 Tripod, AV07–AV12 Machine Data (50 Hz, UL)	FINO1 (1 Hz, 20 Hz, UL)	AV00 (0.2 Hz, UL)
L_{tol}	5	5	5	15	1	0
t_{crit} [s]	1	1	1	1	50, 2.5	250
n_{crit}	0					
r_{crit}	0.97					0.8
w [h]	0.005	0.06	0.06	0.06	3, 0.15	15
m	4					

Abbreviations

BSH	Bundesamt für Seeschifffahrt und Hydrographie/ Federal Maritime and Hydrographic Agency
DNV GL	GL Garrad Hassan Deutschland GmbH (formerly WINDTEST Kaiser-Wilhelm-Koog GmbH) as part of DNV GL
IEC	International Electrotechnical Commission
OFFIS	OFFIS e.V. Institut für Informatik, Oldenburg
RAVE	Research at Alpha Ventus
UL	UL International GmbH (formerly DEWI GmbH or the Deutsches Windenergie-Institut in Wilhelmshaven)