Quality control of raw data during measurement campaigns for wind energy

R. Foreman, N. Hansen, T. Neumann
UL International GmbH, Wilhelmshaven, Germany, r.foreman@dewi.de;

Summary
The assessment and quality control of raw data (pre-processing) can be time consuming, particularly considering the large quantities of data collected during, as well as the many different types of sensors employed in wind-energy measurement campaigns. Here an automated quality-control procedure is described, which employs simple rules for the detection of common problems encountered during data control, such as spikes in the signal. The tests are applicable for structural, meteorological, oceanographic and electrical data, with a minimal specification of the parameters required for each unique dataset. The tests are not designed to be completely independent of human assessment, but rather to aid and accelerate the initial assessment of data quality.

1. Introduction

Disturbances in raw signals from unphysical behaviour, resulting from faults in the measurement system, for example, can affect the results of analyses if not filtered beforehand. However, the user is often abandoned to the task of manually sorting through large quantities of time series before even beginning their analysis. The following procedure aims to accelerate this task.

2. Methodology

2.1 Description of Tests

The quality-control procedure consists of the seven tests summarized in Table 1. The timing, length and flat line tests are relatively formal, checking, respectively, the correctness of the timestamp, the length of data, and whether the sensor is still active. The measurement range, bad resolution and spike tests check whether the signal lies beyond the amplifier range, and for the presence of quantization and outliers, respectively.

Tab. 1: Summary of quality-control procedure

Test Name	Description
Timing	Checks timestamp
Length	Correct length of data
Flat Line	Constant value
Measurement Range	Data outside range
Bad Resolution	Checks for quantization
Spike	Outliers in signal
Visual Inspection	

2.2 Flagging Strategy

The results of each test described in Tab. 1 are translated into a value or ‘flag’ with the particular qualitative meaning as summarized in Tab 2. The first three tests apply only Flags 1 and 4 as they are based only on simple rules. The next three tests apply Flags 1–3 based on a statistic between the raw data and the output data after undergoing the testing. For the spike test, for example, the raw data is despiked, with the flagging based on the correlation between the despiked and raw signal. Badly correlated data (yielding Flag = 3) are then referred to the visual inspection test, which may then give a data classification as good (Flag = 1) or bad (Flag 4) after closer inspection. This procedure reduces the time of manual inspection while minimizing the number of falsely flagged records.

Tab. 2: Description of flagging strategy

Flag	Meaning	Description
1	Good data	Measurements can be used safely
2	Probably good data	User should verify data before use
3	Problems with data	Data probably need to be corrected before use
4	Bad data	Measurements should be rejected.

3. Example

3.1 Spike Test

Figure 1 shows part of a signal affected by a periodic electrical disturbance containing the spike, which the spike test has removed. The flag for the spike test is based on the correlation between the raw and despiked signal. Here, for example, the spike reduces the correlation coefficient to a value < 0.999 and is flagged, but is still useable after despiking.

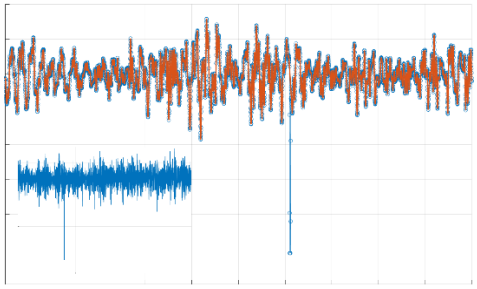

Fig. 1: Example of part of signal affected with a single spike. Inset: the entire time series (30,000 points).