AStrion data validation of non-stationary wind turbine signals

Guanghan Song, Zhong-Yang Li, Pascal Bellemain, Nadine Martin, Corinne Mailhes

HAL Id: hal-01182956

https://hal.science/hal-01182956

AStrion data validation of non-stationary wind turbine signals

G Song, Z-Y Li, P Bellemain, N Martin and C Mailhes

AStrion is an automatic spectrum analyser software, which proposes a new generic and data-driven method without any a priori information on the measured signals. In order to compute some general characteristics and derive the properties of the signal, the first step in this approach is to give some insight into the nature of the signal. This pre-analysis, the so-called data validation, contains a number of tests to reveal some of the properties and characteristics of the data, such as the acquisition validity (the absence of saturation and a posteriori in respect of the sampling theorem), the stationarity (or non-stationarity), the periodicity and the signal-to-noise ratio. Based on these characteristics, the proposed method defines indicators and alarm trigger thresholds and also categorises the signal into three classes, which helps to guide the following spectral analysis. The present paper introduces the four tests of this pre-analysis and the signal categorisation rules. Finally, the proposed approach is validated on a set of wind turbine vibration measurements to demonstrate its applicability for a long-term and continuous monitoring of real-world signals.

1. Introduction

Wind turbines are often installed in remote areas that are hard for technicians to access. Hence, the maintenance of wind turbines is very difficult to carry out. By allowing a predictive maintenance strategy, condition monitoring systems (CMS) are an efficient solution for the maintenance of wind turbines. To track the health of the wind turbine devices, CMS are often installed to alert users of mechanical faults[1]. In general, CMS are based on the acquisition of a set of relevant real-time signals from the sensors, and then signal processing is further applied to deduce health indicators of the system. To realise this system health maintenance, there are two kinds of method: system-driven and data-driven. System-driven methods are more commonly used, while the monitored system's kinematic and specific thresholds are predefined by the users. Different from system-driven methods, data-driven methods do not need any *a priori* information but depend only on the measured process data[2].

In our context, the data-driven CMS are preferred since they do not require any user configuration to set the monitoring system functionality. This type of CMS requires the signal processing algorithm to be standalone and self-configurable using only the acquired data to set up all necessary parameters. Moreover, it has to be reliable in all kinds of situations to liberate the users from unexpected manual maintenance tasks.

AStrion answers this need by elaborating a set of comprehensive spectral analyses. It is capable of deducing reliable mechanical fault indicators thanks to many spectrum analysis modules, which are described in previous publications[3,4]. In this paper, we focus on the angular resampling and data validation modules in AStrion. In general, angular resampling is an optional module to reduce the problem of non-stationarity in wind turbine signals, while data validation provides a pre-analysis of the acquired signals.

Wind turbines often operate in variable environmental conditions, such as the wind speed, the temperature and the azimuth, that result in different non-stationarity patterns. To reduce the influence of the non-stationarity in spectral analysis, we propose an angular resampling method to reduce the non-stationarity caused by a variable rotational speed.

Among all the data validation tests of AStrion, in this paper we focus on four tests to check four properties of wind turbine signals. The first test verifies whether any amplitude saturation exists in the instrumentation process. The second test verifies whether the signal is well sampled. The third test is on the stationarity. Under the hypothesis that the spectral content of the vibration signal of a machine is time-invariant during a limited time length, the time-varying spectral content indicates the non-stationarity of the signal. The fourth test checks whether the signal is periodic, reveals the periodicity of the signal and estimates the signal-to-noise ratio. Combining the results of the four tests, the properties of the signal are highlighted, which are used to set up the following analysis strategy.

In this paper, we first present the methodology of angular resampling and the tests in the data validation module. Then, we demonstrate the accuracy of the proposed method on a wind turbine simulation test-bench and also the applicability of the proposed method for the characterisation of real-world wind turbine signals.

2. Angular resampling

AStrion-A is an angular resampling module in AStrion. The angular resampling aims to transform a time-domain signal in the angular domain with the help of the phase marker signal. In such a way, the non-stationarity caused by the variation of the rotational speed can be eliminated, since the signal sampling is synchronised

with the rotation. Let us note the vibration signal as $s(n)$ and the raw tachometer signal as $I[n]$, where n is the discrete time index: $n \in \mathbb{Z}^+$, $n \leq N$, with N being the number of samples.

2.1 Extraction of the time length of each revolution

The ideal phase marker signal is a noise-free pulse train and the time-of-arrival of each pulse clearly indicates a revolution. However, raw phase marker signals can be recorded in a non-impulse format, sometimes with an additive noise. To identify the time length of each revolution from a raw signal, the first step is to take only the positive valued part of the derivative of the raw tachometer signal, $I[n]$, to generate a pseudo-impulse format:

$$I_d\left[n\right] = I\left[n+1\right] - I\left[n\right], \text{ for } 1 \leq n \leq N \text{ ....} (1)$$

Then, three thresholds are set up:

$$\lambda_1 = \sqrt{2 * Var\left(I_d\left[n\right]\right)} + Mean\left\{I_d\left[n\right]\right\}, \quad \lambda_2 = \frac{\lambda_1 + \lambda_3}{2}, \quad \lambda_3 = \frac{Max\left(I_d\left[n\right]\right)}{\sqrt{2}}$$

The pseudo-impulse tachometer signal is finally transformed to the ideal format of a noise-free impulse train:

$$I_\delta\left[n\right] = \sum_{i=1}^{N_\delta} \delta\left(n - n_i\right) \text{................................} (2)$$

where $\delta(n)$ is a Kronecker delta, N_δ is the number of pulses and n_i is the discrete time index of the ith sample. The transformation is realised by the following pseudo-code:

Repeat the following steps for each λ in $\{\lambda_1, \lambda_2, \lambda_3\}$.

- Obtain $I_{\delta,\lambda}\left[n\right]$ by thresholding $I_\delta\left[n\right]$ with λ:

$$I_{\delta,\lambda}\left[n\right] = \begin{cases} I_d\left[n\right] & I_d\left[n\right] > \lambda \\ 0 & I_d\left[n\right] \leq \lambda \end{cases} \text{...................} (3)$$

- Localise the pulse, n_i, by finding the maximum of the λ-th cluster of non-zero points, where the number of consecutive zero-valued points is equal to or smaller than four to reduce the false identification of pulses.
- Calculate the time length of each revolution as the interval between two adjacent impulses, such as $T_\lambda\left[i\right] = n_i - n_{(i-1)}$.
- Calculate the coefficient of variation $cv_\lambda = \dfrac{std\left\{T_\lambda\left[i\right]\right\}}{mean\left\{T_\lambda\left[i\right]\right\}}$.
- Update $T[i] = T_\lambda\left[i\right]$ if cv_T is lower than the value of the previous iteration.

Finally, the revolution time-length vector is obtained as $T[i]$ for $1 \leq i \leq N_\lambda$.

2.2 Time-length adjustment and speed calculation

In raw phase-marker signals, the identification of the revolutions is disturbed by the noise. The smooth, varying non-impulse waveforms make the identification even more difficult. These types of disturbance often result in false measurements and outliers in the time-length vector. Therefore, the time-length vector has to be adjusted to compute the correct rotational speed. Firstly, identify the outliers in $T[i]$ by calculating the difference between $T[i]$ and a reference time-length vector, $\overline{T}\left[i\right]$:

$$\tilde{T}\left[i\right] = T\left[i\right] - \overline{T}\left[i\right] \text{ with } \overline{T}\left[i\right] = Median\left\{T\left[i-1\right], T\left[i\right], T\left[i+1\right]\right\} \text{ .... } (4)$$

since we assume that the rotational speed is stable over three consecutive revolutions. The outliers of $T[i]$ can be found as the outliers of $\tilde{T}\left[i\right]$ by the following procedure:

Initialise the outlier set $X = \varphi$; carry out the following iteration:

- Calculate the Z-score of $\overline{T}\left[i\right]$ as $zs\left[i\right] = \dfrac{\tilde{T}\left[i\right] - Mean\left\{\tilde{T}\left[i\right]\right\}}{Std\left\{\tilde{T}\left[i\right]\right\}}$.
- Localise i_{max}, the maximum of $zs[i]$, and exclude it by $zs[i_{max}] = Mean\{zs[i]\}$.
- Calculate the coefficient of the variation of $zs[i]$ as $cv_z s = \dfrac{Std\left\{zs_{rem}\right\}}{Mean\left\{zs_{rem}\right\}}$.
- $X = X \cup i_{max}$; Continue iterating if $cv_{zs} > 0.01$.

Then, calculate the phase shift $\varphi[i]$ and the angular speed $v[i]$ in each revolution i as:

$$\varphi\left[i\right] = \begin{cases} \dfrac{2\pi}{K}\dfrac{T\left[i\right]}{\overline{T}\left[i\right]} & i \in X \\ \dfrac{2\pi}{K} & \text{otherwise} \end{cases}, \quad v\left[i\right] = \dfrac{\varphi\left[i\right]}{2\pi}T\left[i\right]F_{S_I} \text{ ... } (5)$$

where K is the number of pulses emitted by the tachometer in every revolution and F_{s_I} is the tachometer sampling frequency. The revolution-wise speed is interpolated to each time sample, n, to calculate the instantaneous rotational speed:

$$v_{inst}\left[n\right] = Interp\left(v\left[i\right]; i \rightarrow n\right) \text{........................} (6)$$

2.3 Phase calculation and resampling

Calculate the instantaneous phase $\varphi_{inst}[n]$ and the angular-resampled instantaneous phase $\theta[n]$ as:

$$\varphi_{inst}\left[n\right] = \varphi_0 + \frac{2\pi}{Fs_{vib}}\sum_{m=1}^{n} v_{inst}\left[m\right], \quad \theta\left[n\right] = \varphi_0 + \frac{\varphi_{inst}\left[N\right] - \varphi_0}{N}n \text{ ... } (7)$$

where $\varphi_0 = \dfrac{2\pi}{K} - \varphi\left[1\right]$ is the initial phase. The signal resampled in the angular domain is obtained by interpolating the vibration signal $x[n]$ from $v[n]$ to $\theta[n]$:

$$x_{res}\left[n\right] = Interp\left(x\left[n\right]; \varphi_{inst}\left[n\right] \rightarrow \theta\left[n\right]\right) \text{.............} (8)$$

3. Data validation methods

In order to estimate basic global and general properties regarding the signal, the first step in this data-driven approach is to collect fundamental information about the input signal nature. As mentioned in Section 1, a number of tests based on spectral analysis from the Fourier type estimators are carried out on the input signal to reveal its nature. This pre-analysis consists of successive and independent steps. One or more criteria are taken into account without any *a priori* information of the signal for each step. With respect to the amplitude saturation, the sampling validity, the non-stationarity and the periodicity, four steps are described below. This function is realised by module AStrion-D.

3.1 Amplitude saturation test[5]

This is a simple test to provide an indication regarding the amplitude saturation, which is performed by comparing two consecutive signal samples to the maximum and minimum of the signal. If there are two consecutive samples equal to either the maximum or the minimum, the signal is considered to be saturated.

3.2 Shannon sampling test[5]

The aim of this test is to detect potential problems with the signal recovery from its spectrum due to aliasing. In particular, three tests are performed: a test to verify *a posteriori* if the signal is properly sampled (by observing the spectrum as we approach the Shannon frequency); a test to determine the last significant frequency component (by observing the logarithm of the spectrum); and a test to confirm the presence of an anti-aliasing filter (by observing the noise estimation). To evaluate the performance, three criteria are carried out, such as linear power frequency, decibel power frequency and noise power frequency.

3.3 Non-stationarity tests

The definition of a stationary process can explain that there exists no unique efficient test. An ergodic random process is completely stationary if its probability density function remains unaltered under a time shift. This definition is a severe requirement and should be relaxed. A time-frequency test was proposed. Then, using the result of this time-frequency test, a second test based on the signal spectrum only will confirm, refute or eventually detect other types of non-stationarities[6].

The first time-frequency-based detection test is to verify whether the analysed signal contains non-stationarities in terms of a variation of its spectral contents *versus* time. The second time-based detection test provides a measure of stationarity of the signal in a statistical sense. A process is statistically stationary if it is characterised by time-independent moments. The second, third and fourth moments are calculated for N consecutive non-overlapping segments of the input signal. A significant variation of these quantities, detected by thresholding the cumulative sum, indicates that the signal is non-stationary.

3.4 Periodicity tests

These tests validate the assumption that the considered signal, corrupted by noise, is periodic in time. It is performed by using the autocorrelation function of the signal, which preserves the periodicity property. Two criteria are taken into account. For each one, a coefficient between 0% and 100% is calculated in order to quantify the hypothesis confidence that the underlying signal is a periodic signal embedded in white Gaussian noise[7].

The first criterion of fundamental periodicity is estimated as a median value of distances between consecutive maxima of the autocorrelation function. The non-biased autocorrelation estimation is used; maxima are determined considering the estimation variance. The second one is the dominant-power periodicity test, which does not necessarily coincide with a fundamental one. It is estimated as a median value of distances between consecutive minima of the smoothed autocorrelation function.

3.5 Strategy

The following identification function is driven by the data validation results. The data validation does not reveal the spectral characteristics of the signal, it gives us a clue about the nature of the analysed signal. Rough conclusions regarding this nature, and conclusions regarding the periodicity of the signal, the correlation support and the signal-to-noise ratio (SNR), define three possible strategies that are implemented in the identification block.

4. Results of characterisation of real-world signals

The proposed angular resampling and data validation methods have been tested on real-world signals from two datasets. One records the signals over a wind turbine simulation test-bench[3]. The other, named Arfons, is composed of real-world wind turbine signals.

4.1 Data validation on test-bench data

The test-bench developed in the KAStrion project has been specified to design a wind turbine on a smaller scale. The experimental platform is monitored by several sensors, including accelerometers, thermocouples and torquemeters, etc. To test the proposed method, we focus on an accelerometer, which is installed on the main bearing with the direction (+y). Figure 1 illustrates the position and the direction of the accelerometer.

Figure 1. The wind turbine test-bench. The orange ellipsis circles zoom into the main bearing and its loading unit, while the green arrows represent the three wired accelerometers

Data validation was applied to the signals acquired from the main bearing, which has been fully damaged up to totally stopping normal operation. In total, 20 signals have been recorded. The first 75 s of each signal with a sampling frequency of 3.9 kHz were calculated. Part of the surveillance of the results is presented in Figure 2. The results of the four tests in AStrion-D are detailed hereinafter:

- **Time saturation test:** The input signals passed the saturation test normally. We consider the input signal has no time saturation.
- **Shannon sampling test:** The decibel spectrum test detects the aliasing of strong power frequency components close to the Shannon frequency. Figure 3(a) presents the power spectral density map of the signal acquired at an operating time of 85.97 h. The power is mainly concentrated at low frequencies. However, the decibel and noise power maximum frequency estimators detected high-power peaks appearing in high frequencies (16 kHz and 19 kHz), which is unusual. These high-power peaks in high frequencies appear all the time, which may be due to an instrumentation fault.
- **Non-stationarity test:** The experimental condition of the test-bench is stationary, hence the signals should not have a stationarity problem. In Figure 4, we can see that the signal has a non-stationarity problem around an operating time of 168.57 h. The details of time-frequency, non-stationarity detection and its variance are presented in Figure 5. The non-stationarity test detects a hardware breakdown during the acquisition of the signal. After six hours, signals from this sensor
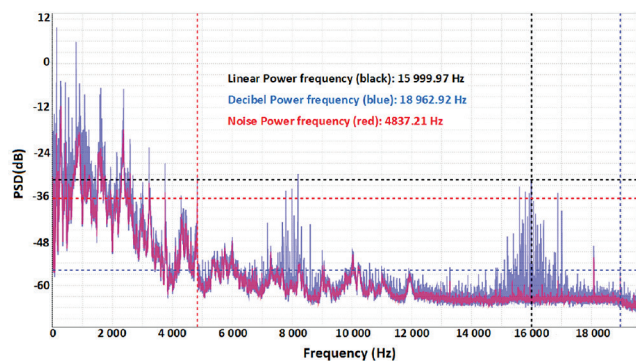
Data validation								
Time saturation test								
Shannon sampling test (LN, dB, N)	L dB	L dB	dB	dB	dB		dB	dB	dB
Test LN Linear power frequency/order	19134	19134	17010	17010	16000	3224	16000	16000	16000
Test dB Decibel power frequency/order	19513	19514	18343	17892	18685	16000	19070	18651	19396
Test N Noise power frequency/order	6545	6546	6830	6830	3763	2490	3763	4837	3763
Time-frequency detection (non-stat. rate)	0%	0%	1%	11%	0%	99%	0%	0%	0%
Time detection indicator	4 to 16	6 to 20	3 to 24	5 to 24	5 to 21	2 to 8	3 to 20	3 to 20	2 to 22
Fundamental periodicity rate	26,67%	11,9%	40%	40%	5,298%	54,8%	18,11%	15,38%	23,97%
Fundamental periodicity frequency/order	1,46	1,101	0,167	2,239	2,24	781,3	32,85	0,462	34,27
Fundamental periodicity rate SNR (dB)	0	0	0	0	0	-10,69	0	0	0
Dominant-power periodicity rate	71,21%	63,24%	44,09%	74,46%	100%	100%	100%	99,66%	100%
Dominant-power periodicity frequency/order	783	429	418,2	434,1	788,4	788,5	788,4	788,5	788,4
Correlation support	100%	100%	100%	100%	100%	63,08%	100%	100%	100%
Nature of the signal	S QP SN	S SN	NS SN	NS QP SN	S QP SN	NS QP SN	S QP SN	S QP SN	S QP SN

Figure 2. The behaviour of AStrion-D over time. The green colour is OK and the red colour is ALARM. S is for stationary, QP is for quasi-periodic and SN is for significant noise

(a)

(b)

Figure 3. Power spectral density (PSD) of the signal acquired from: (a) test-bench data at an operating time of 85.97 h; and (b) Arfons data of the 77th signal

Figure 4. Surveillance of the non-stationarity test of the accelerometer (+y) (main bearing of the test-bench) from 0.96 h to 189.85 h of operation

are unrecorded because of a cable fault.

- **Periodicity test:** The indicators such as fundamental periodicity frequency and correlation support of the signal length also indicate the hardware breakdown at an operating time of 168.57 h (shown in Figure 6). In Figure 6(a), the fundamental frequency is detected at a high value for the first three signals,
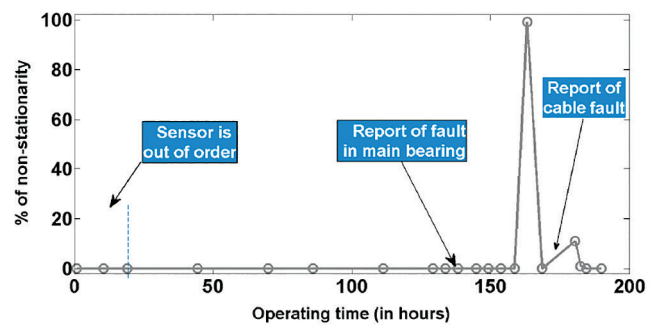
which are pure noise. For the following signals, the fundamental frequency is stable at around 2.5 Hz, until an operating time of 168.57 h with increments to 781.3 Hz. In Figure 6(b), the higher correlation support value indicates a larger periodic percentage in the whole signal, except for the first three noise signals (not periodic at all), where only the signal at the operating time of 168.57 h decreased to 63% from 100%.

4.2 Results on Arfons data

Arfons data came courtesy of Valorem, France, and are acquired from the on-site installation of sensors in a wind farm. Two wind turbines of the size described in Figure 7(a) have been equipped with the identical hardware and software configuration of the developed system. Continuous monitoring and signal acquisition is carried out using vibration, current and voltage sensors. Figure 7 shows the size of the wind turbine and the location of the sensors. In the following calculation, 77 signals from the accelerometer mounted at the rare end of the generator WT6 (A6 in Figure 7(b)) are selected between 20 December 2014 to 7 January 2015. The selected signals are of 10 s, sampled at 25 kHz under normal operation states with a mean rotational speed between 1600 r/min and 1800 r/min.
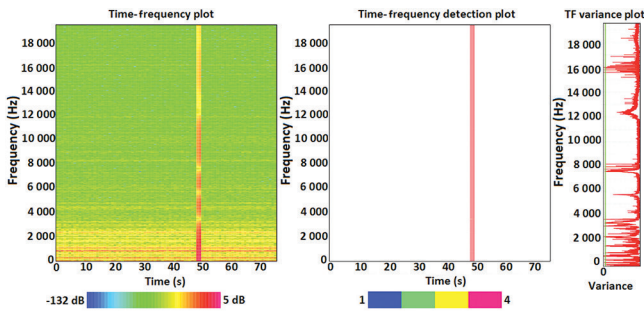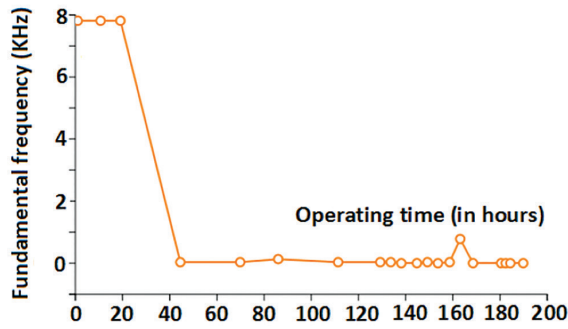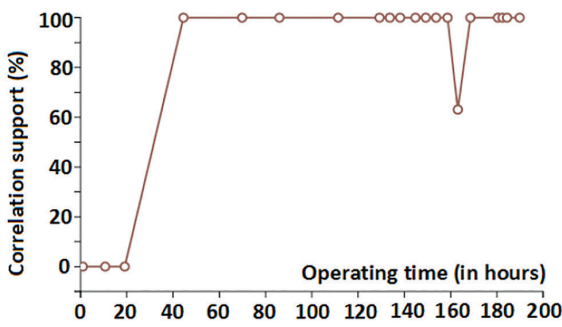
Figure 5. Time-frequency spectrogram (left), non-stationarity detection (middle) and its variance (right) of the signal at an operating time of 168.57 h
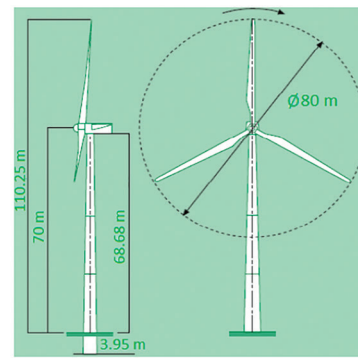

Figure 6. Surveillance of the periodicity tests of the accelerometer (+y) (main bearing of the test bench) from 0.96 h to 189.85 h of operation: (a) detected fundamental frequency; (b) correlation support of the signal length
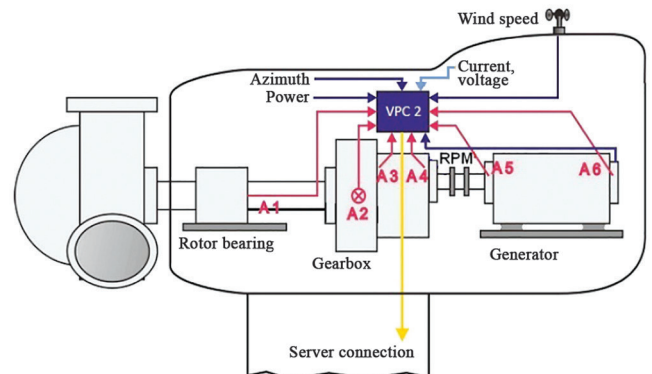
Since the rotational speed of the wind turbine is variable, the speed-related spectral content is located on frequency trajectories varying proportionally to the rotational speed. This is the main source of non-stationarity of a signal. The angular resampling is an efficient solution to cancel this type of non-stationarity, as Figure 8 shows.

4.2.1 Angular resampling

In Figure 8, the red curves in the detection plot (on the middle of each Figure) mark the frequency trajectories where non-stationarity is detected either in the amplitude or in the frequency. At the medium- and low-frequency range (lower than 7700 Hz) of Figure 8(a), the non-stationarity trajectories correspond to the speed-related harmonic series. In Figure 8(b), the disappearance of these non-stationary trajectories (lower than 280 orders) proves that the time-varying frequency becomes constant due to the angular resampling. However, on both subfigures, the high-
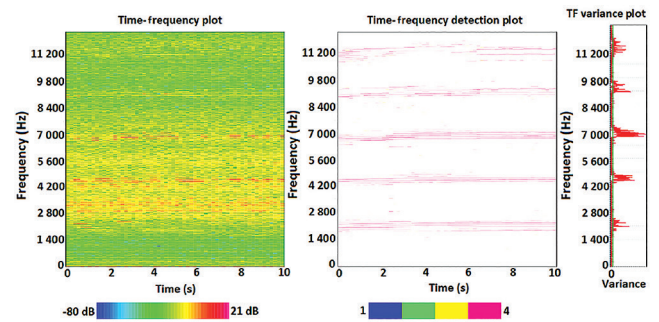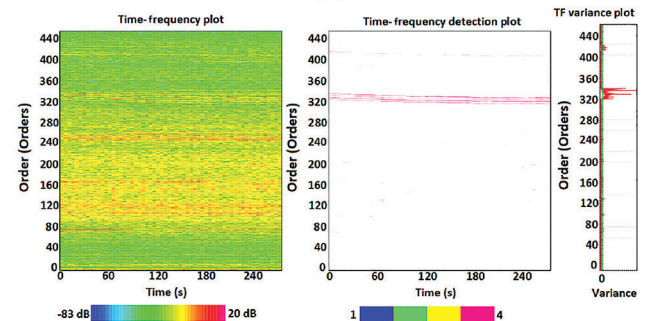

(a)

(b)

Figure 7. General information of the wind turbines under study: (a) dimensions of the wind turbines; (b) location of the installed sensors

(a)

(b)

Figure 8. The spectrogram and the non-stationarity detection results of the same signal (75th measurement) without (a) and with (b) the angular resampling (sampled at 911.7 orders). The time-frequency spectrogram (left), non-stationarity detection (middle) and its variance (right)

frequency trajectories (higher than 7700 Hz in Figure 8(a) and larger than 280 orders in Figure 8(b)) cannot be eliminated by the angular resampling, because the non-stationarity is not produced by the variation of the rotational speed.

The details in the time-frequency domain can be found in Figure 9, where the presentation of the results of Figure 8 are zoomed at the medium- and at the low-frequency range (lower than 2500 Hz and lower than 92 orders).
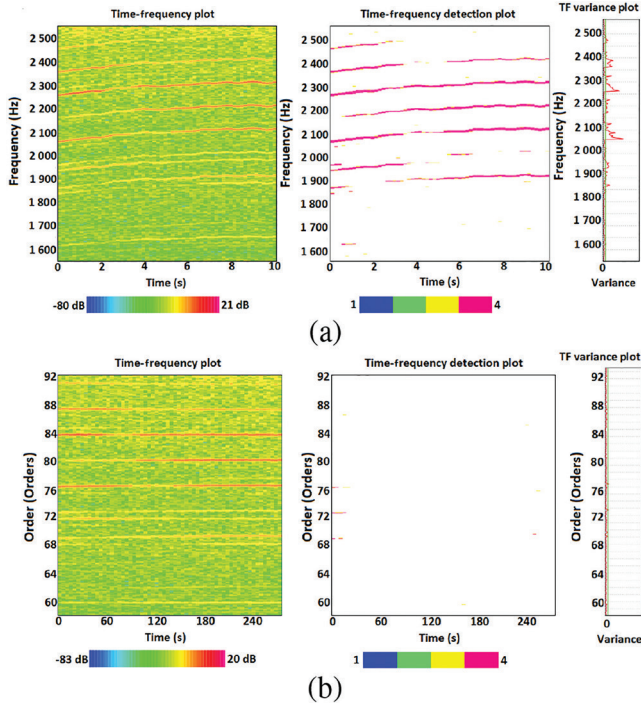

Figure 9. A zoomed presentation of the results of Figure 8 at the medium- and the low-frequency range (lower than 2500 Hz and lower than 92 orders)

Compared to the spectrogram of the original signal, which presents a series of time-varying frequency trajectories, that of the resampled signal contains a group of constant-frequency components. The non-stationarity trajectories are almost eliminated in the resampled signal. Accordingly, a significantly lower non-stationarity rate can be found in the resampled signal (5%, alert level) with respect to the original signal (19%, alarm level).

On the whole dataset of 75 signals, the effect of the angular resampling in the reduction of the non-stationarity is evident, as shown in Figure 10. It successfully reduced the non-stationarity
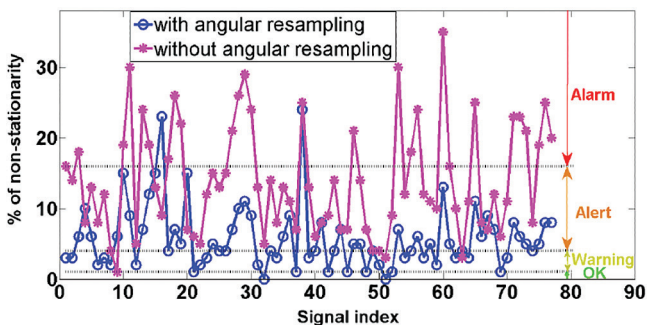

Figure 10. The non-stationarity rate with and without angular resampling on 75 signals of accelerometer A6 of the Arfons wind farm

rate by a maximum of 20%. The non-stationarity rate is a log-scale index, hence the gain of using the angular resampling is enormous for these signals. However, there are signals which have higher or equivalent non-stationarity rates after angular resampling, for example the 16th and the 38th signals, since the non-stationarity is not related to the rotational speed. In this case, the angular resampling shows little benefit.

4.2.2 Data validation

- **Amplitude saturation test:** The input signals passed the saturation test normally. We consider the input signal has no time saturation.
- **Shannon sampling test:** The decibel spectrum test detects the aliasing of strong power frequency components close to the Shannon frequency. Figure 3(b) presents the power spectral density map of the 75th signal. Unlike test-bench data, this power is mainly concentrated in the high order. The decibel power maximum frequency estimator detects a high-power peak appearing in the 427.14 order, which is unusual. These high-power peaks in the high order appear all the time. It may be due to the electromagnetic disturbance or the resonance of other structures.
- **Non-stationarity test:** In wind turbine signals, the non-stationarity problem frequently appears. To reduce the non-stationarity influence on the following calculation, we proposed angular resampling (described in Section 4.2.1), which indeed reduces the non-stationarity percentage.
- **Periodicity test:** Figure 11 presents the surveillance of the indicators: the fundamental frequency and the correlation support. In (a), we can normally detect the fundamental frequency around the 82 order with some outliers. In (b), the correlation support value is normally 100% with some outliers.
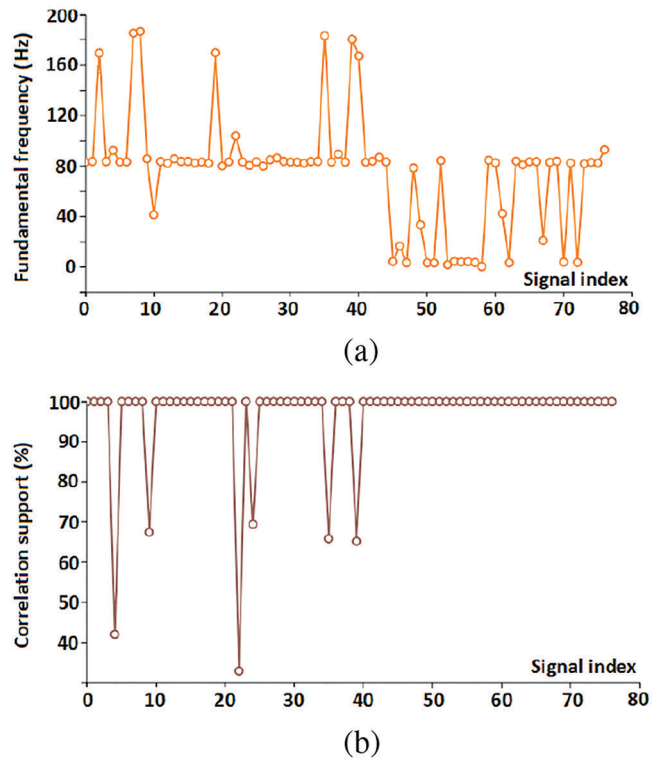

Figure 11. Surveillance of the periodicity tests on 77 signals of accelerometer A6: (a) detected fundamental frequency; (b) correlation support

5. Conclusions

In this paper, we present two methods to validate some necessary properties of the acquired signal. The angular resampling method is able to retrieve the rotation speed from non-impulse or noisy phase marker signals, and it is then able to correct the speed measurement faults by calculating the true phase shift. The data validation step consists of several tests based on spectral analysis and the autocorrelation function to collect fundamental information about the nature of the input signal. The proposed methods are part of the first two modules in an automatic spectrum analyser called AStrion, the core of a data-driven CMS.

The results show that the proposed angular resampling method efficiently reduces the non-stationarity rate on real-world wind turbine signals. Also, the data validation step can help to detect the signals that have acquisition errors. These corrupted signals are not suitable for the following analysis. Moreover, the surveillance of the indicators of the data validation may indicate an early problem with acquisition due to a hardware breakdown.

References

1. P F G Márquez, A M Tobias, M J P Pérez and M Papaelias, 'Condition monitoring of wind turbines: techniques and methods', Renewable Energy, Vol 46, pp 169-178, 2012.
2. S Yin, G Wang and H R Karimi, 'Data-driven design of robust fault detection system for wind turbines', Mechatronics, Vol 24, pp 298-306, 2014.
3. T Gerber, N Martin and C Mailhes, 'Monitoring based on time-frequency tracking of estimated harmonic series and modulation sidebands', CMMNO 2014, Lyon, France, December 2014.
4. Z-Y Li, T Gerber, M Firla, P Bellemain, N Martin and C Mailhes, 'AStrion strategy: from acquisition to diagnosis. Application to wind turbine monitoring', CM2015/MFPT2015, Oxford, 2015.
5. D Matthieu, 'Une stratégie pour l'interprétation en analyse spectrale', PhD thesis, 1999.
6. N Martin and C Mailhes, 'A non-stationary index resulting from time and frequency domains', CM2009/MFPT2009, Dublin, 2009.
7. N Martin and C Mailhes, 'About periodicity and signal-to-noise ratio – the strength of the autocorrelation function', CM2010/MFPT2010, UK, 2010.