Fault detection and isolation of bearings in a drive reducer of a hot steel rolling mill

Marcello Farina [a,*], Emanuele Osto [b], Andrea Perizzato [a], Luigi Piroddi [a], Riccardo Scattolini [a]

[a] *Politecnico di Milano, Dipartimento di Elettronica, Informazione e Bioingegneria Via Ponzio 34/5, 20133 Milano, Italy*
[b] *Primetals Technologies Italy Srl*

Received 28 October 2014
Accepted 2 February 2015

1. Introduction

In plants involving rotating machinery, defects in bearings can crucially affect the process quality (Randall, 2011). For this reason, the early and automatized identification of faulty bearing conditions is of great importance in industrial practice (Ericsson et al., 2005), fostering research in fault detection and isolation.

Fault detection methods can be roughly classified into two wide categories: model-based methods and data-driven (also denoted history-based) methods. The former require the availability of a physical model of the system under analysis, the derivation of which is a complex and time consuming task in the case when the dynamics of bearings is concerned. Data-driven methods are quite popular in the literature (see, *e.g.*, Randall, 2011; Zhang et al., 2011) because they do not require any specific *a priori* knowledge on the system characteristics and only rely on the processing of available data measurements and statistical techniques. A great variety of methods falling in the latter category have been proposed in the literature, ranging from time-domain to frequency-domain techniques, including mixed time–frequency methods.

In this work we propose a two-step scheme, relying on complementary data-driven techniques, for fault detection and isolation for a drive reducer in a hot steel rolling mill. Specifically, a preli-minary fault detection phase is carried out based on a simple and computationally lightweight time-domain analysis algorithm. Secondly, a more complex and computationally intensive method is proposed, based on frequency-domain analysis, targeting a more precise identification of the fault characteristics. Although there are several measurable system variables that can provide information on an incipient bearing fault (*e.g.*, temperature, vibration and acoustic noise), both the proposed procedures essentially require vibration measurements solely. These signals are easily collected during system operation in the considered test rig.

The preliminary fault detection method uses Statistical Process Control (SPC) techniques based on Hotelling's T^2 distance (Montgomery, 2009), computed from the available vibration signals. SPC is a classical statistical methodology for monitoring the process operation and assess if nominal safety conditions are preserved. More precisely, multivariate statistical analysis is used to define a multi-dimensional confidence region for a set of process-related variables, which can then be used to determine if the considered set of variables remains in the safe region. Multivariate statistical techniques are extremely appealing to their simplicity, and are often employed for process monitoring, fault detection and diagnostic tasks on process plants, and specifically for the detection of gearbox defects (Baydar, Chen, Ball, & Kruger, 2001; Ge, Kruger, Lamont, Xie, & Song, 2010) A completely automatized procedure is proposed to process the raw signals and perform the fault detection. Briefly, several vibration signals are first collected performing differ-ent experiments on the test rig in normal operating conditions. Signal portions characterized by sufficient signal level and quasi-stationarity are isolated in the raw data. Then, these signal portions are further divided into subsequent short data blocks, each of which is synthetically characterized by a set of aggregate features. A reference model is formed based on this aggregate characterization of

* Corresponding author. Tel.: +39 02 2399 3539, fax: +39 02 2399 3412.
E-mail addresses: marcello.farina@polimi.it (M. Farina), emanuele.osto@siemens.com (E. Osto), luigi.piroddi@polimi.it (L. Piroddi), riccardo.scattolini@polimi.it (R. Scattolini).

the vibration signals. Faulty conditions are finally identified by monitoring if the data collected during normal plant operation display significant deviations (in a statistical sense) from the refere-nce model.

Since multivariate analysis does not provide any additional information regarding the cause or location of the detected bearing fault, the subsequent step aims to identify the frequency at which a fault occurs and the type of fault, such as ball or cage damage or inner/outer race defect. Fault frequencies are particularly important for fault isolation, since direct correspondences between specific faults and their characteristic frequencies are *a priori* known (Randall, 2011). This motivates the introduction of a second frequency-domain procedure to complement the outcome of the time-domain one described above. Notice that standard spectral analysis is not particularly effective in this case, since the amplitude of the impul-sive type signals associated to defects of the bearings is typically much smaller than that of the signal measured in the absence of fault. Thus, it becomes necessary to employ specific frequency domain techniques that enable to discriminate the contributions associated to faults from the characteristic frequencies of the system in the spectra of the vibration signal. For this purpose, one can resort to the envelope analysis, a well-established technique, see Randall and Antoni (2011), widely employed in the diagnosis of rotation machines. This analysis requires first that the data be appropriately filtered to emphasize the fault with respect to other components, so as to facilitate the fault detection. The design of such filter is here based on the analysis of the spectral kurtosis (Randall & Antoni, 2006; Sawalhi & Randall, 2004). Another automatized procedure has been developed to process the measured signals using the explained envelope analysis. Since this procedure is computationally much more intensive than the previous one based on multivariate analysis, it provides a complementary tool, which can be employed when the former suggests the possibility of a fault, with the aim of providing further evidence and details on the fault.

A major improvement to standard envelope analysis is here introduced, to cope with the case of time-varying shaft speed typically encountered in applications as the one considered. Indeed, since fault frequencies are linearly related to the shaft angular frequency, the variability of the shaft speed blurs the periodic pattern associated to the defect, complicating its detection. For this reason, we here operate on vibration signals represented as a function of the motor shaft angle rather than of time, since in the "angle domain" the fault periodicity can be clearly detected. The two fault detection procedures have been tested using several real data measurements obtained during actual plant operation and simulating the insurgence of defects using standard bearing defect models. The sensitivity of the procedures has been analyzed with respect to amplitude and frequency of the faults. It is shown that the proposed approach (i) can be effective in detecting and isolating incipient local faults at an early stage and (ii) in view of its robustness and reliability features, it is particularly suitable for industrial deployment.

The rest of the paper is structured as follows. The test rig is described in Section 2, together with the measurement system. Section 3 illustrates the first procedure, based on condition monitoring for the detection of bearing-related faults, while Section 4 describes the second procedure, which employs envelope analysis for the frequency-domain characterization of a detected fault. Finally, Section 5 illustrates the application of the presented procedures.

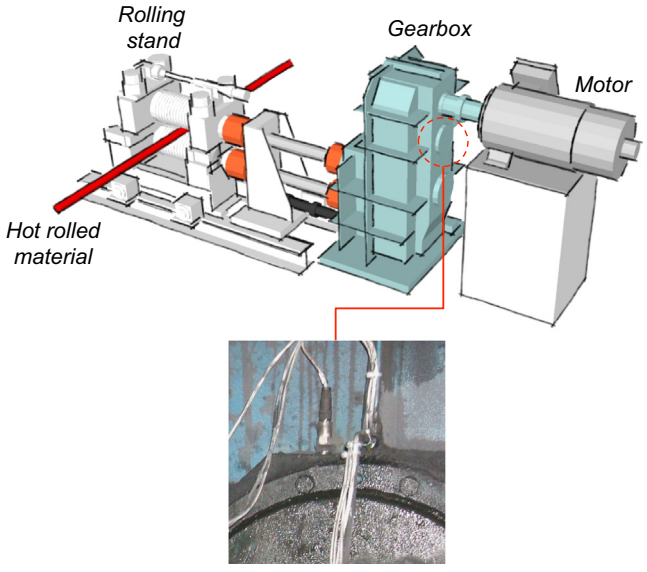

Fig. 1. 3D rendering of the reversible blooming mill to be studied and actual image of the accelerometer.

2. Device description and problem setting

2.1. The drive reducer

The system to be studied consists of a gear reducer serving a reversible blooming mill, represented in Fig. 1.

The hot steel is rolled in multiple runs between the stand's rolling rolls, and at each run the rolled section is reduced and the bar elongates. The turning direction of the cinematic chain is inverted at each run through a deceleration, stop, inversion and acceleration sequence. Seven accelerometers have been installed on the gear reducer to sense the vibration signals on a target group of bearings. The described gearbox application is one of the countless arrange-ments that can be found in rolling mills. A data-driven approach is more suitable to master this complexity, the development of a physical model for each drive arrangement's typology being impractical and anti-economic. As additional benefit, the proposed statistical approach offers a unique fingerprint of the gearbox under observation not replicable with even the most sophisticated physical model.

The objective is to identify any abnormal operating condition of the bearings and possibly isolate the faulty element. Such analysis will be performed using batches of data corresponding to the 4 runs of each steel bar. Note that this does not imply a strong performance limitation compared to real-time analysis, since the bearing degradation is a slow process compared to the duration of each individual batch processing (which typically takes less than a minute).

2.2. Available signals and data features

The available measurements are:

- Seven vibration signals measured by seven accelerometers attached to the shell of the drive reducer. Each accelerometer is placed in the proximity of a bearing, so that it can be assumed that each vibration signal has an essentially one-to-one correspondence with a specific bearing.
- The current absorbed by the motor driving the mill.
- The angular velocity of the motor.

An example of typical vibration and current signals measured in the described plant is reported in Fig. 2.

Several data sets have been recorded in nominal conditions (with no bearing defects detected), accounting for different process runs (bars can have slightly different size and steel quality).

A preliminary frequency analysis (see Fig. 3) reveals several sharp peaks in the amplitude spectra (*e.g.* at 300 Hz, and occasionally at the

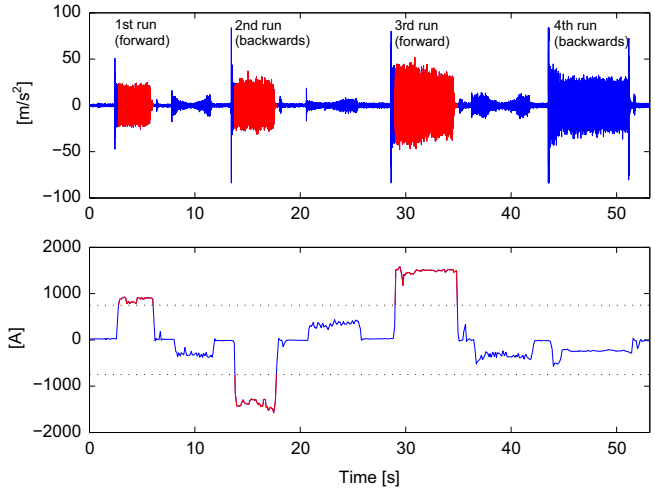

Fig. 2. Vibration (top) and motor current absorption (bottom) for an entire bar processing task (segmented portions in red, see Section 3.1.1). (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

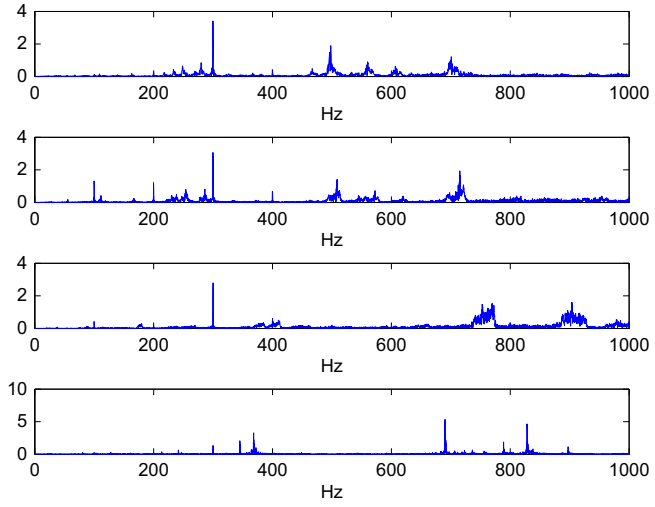

Fig. 3. Spectra of the vibration signals from run 1 (top) to 4 (bottom).

submultiple frequencies) that are not easily explained in terms of the rotation speed of mechanical elements, nor can be associated to the characteristic frequencies associated to faults of the bearings. Most (qualitative) similarities are found between the spectra relative to the first and second sub-periods, as well as the third and fourth, notwithstanding the fact that the runs are executed in alternate verses. In particular, a consistent presence of peaks in the range 500–700 Hz is observed in the first two runs, while the last two display peaks in the range 700–900 Hz. This observation suggests that the absolute value of the rotation speed (or the thickness of the bar), rather than its sign, could affect the characteristics of the measured signals. Finally, the fourth run appears to be radically different regarding the form of the spectrum, and displays clearer peaks.

3. Fault detection by means of condition monitoring

3.1. Data pre-processing

As discussed, the processing of a steel bar consists of 4 runs, where the bar goes back and forth twice into the machine. Consequently, the vibration signals display 4 consecutive portions of significant signal interleaved by short pauses (see Fig. 2). Since the mentioned multivariate statistical methods can only be applied to stationary signals, our first task has been to develop a robust and automatic technique to extract sub-portions of the data that can be regarded as sufficiently stationary for processing and that, at the same time, convey sufficient informa-tion on the bearings behavior. For this purpose we apply the following two-step segmentation procedure to the available data sets:

1. Identify each single run of a specific bar processing.
2. For each run, extract the stationary part of each vibration signal.

In the following, we present two simple algorithms to achieve the desired segmentation.

3.1.1. Individual run identification

To clearly detect the span of each run we can exploit the electrical current absorption signal, which displays high values in correspondence to each run. Based on this, we select the data-portions where the absolute value of the current is higher than some given thresholds. The latter are determined as $i_{pos} = \beta i_{max}$ and $i_{neg} = \beta i_{min}$, for positive and negative current values, respec-tively. Here, $i_{max} > 0$ and $i_{min} < 0$ are the maximum and minimum values assumed by the current signal in a single processing, respectively, and $0 < \beta < 1$ is a user-defined constant. The closer β is to 1, the more likely it is to miss some runs. Conversely, the closer it is to 0, the more likely it is to include spurious data, that do not actually correspond to any run. Therefore, intermediate values of β are expected to provide better results. Fig. 2 displays a typical run segmentation based on a $\beta = 0.5$ thresholding. Notice that the fourth run is not identified at all, due to its much lower current signal level. However, the three preceding runs convey enough information to detect a fault. The value $\beta = 0.5$ has been used in the subsequent analysis.

3.1.2. Identification of the stationary portions of the vibration signals

The measured vibration signals display a non-stationary trend in both the initial and final part of each run, associated to the engagement and disengagement phases of the bar being machined. In particular, the acceleration amplitude can vary significantly in the course of a single run, especially at the extremes. Stated otherwise, the signal variance is not constant, and a stationarity test has to be employed to identify signal portions that can be approxim-ately assumed stationary. The most widely used stationarity tests (*e.g.*, the F-test) are suitable only for Gaussian distributed random variables. We here instead employ the Levene's test (Gastwirth, Gel, & Miao, 2009), which is based on the comparison of the variance of adjacent signal series and is largely insensitive to the non-Gaussianity of the data. A single run acceleration signal Y, consisting of N samples, is divided into k subsequences, each composed of N_i samples. The null hypothesis

$$H_0 : \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2,$$

is tested against the alternative one:

$$H_1 : \exists (i, j) \text{ s.t. } \sigma_i^2 \neq \sigma_j^2.$$

where σ_i^2 denotes the variance of the ith subsequence. Denote with:

- y_{ij} the jth observation of the ith subsequence;
- $\bar{y}_i$ the mean of the observations of the ith subsequence;
- $D_{ij} = |y_{ij} - \bar{y}_i|$ the absolute deviation of the jth observation from the mean of the ith subsequence;

- $\overline{D}_i = (1/N_i)\sum_{j=1}^{N_i} D_{ij}$ the average of the deviations of the ith subsequence;
- $\overline{D} = (1/N)\sum_{i=1}^{k}\sum_{j=1}^{N_i} D_{ij}$ the average of all the deviations.

Levene's test statistic is defined as

$$W = \frac{N-k}{k-1}\frac{\sum_{i=1}^{k} N_i(\overline{D}_i - \overline{D})^2}{\sum_{i=1}^{k}\sum_{j=1}^{N_i}(D_{ij}-\overline{D}_i)^2}$$

Given a significance level α_L, Levene's test rejects the null hypothesis if $W > F_{\alpha_L}(k-1, N-k)$, where $F_{\alpha_L}(k-1, N-k)$ is the upper critical value of the F distribution with $k-1$ and $N-k$ degrees of freedom at a significance level α_L. The value of α_L typically lies in the range [0.01, 0.05]. Algorithm 1 provides the main steps for the application of Levene's test. Algorithm 1 starts from the center subsequence (*i.e.*, the farthest one from the beginning and end of the run, where the largest variance excursions are generally found) and expands the useful portion of the signal as long as the adjacent subsequences have statistically indistinguishable variance. If, at the end of the procedure $L = R = 0$, then there does not exist any stationary part in the considered signal. Otherwise, the stationary part of the signal is defined by the consecutive subsequences from $C-L$ to $C+R$. Fig. 4 illustrates the outcome of the proposed procedure.

Algorithm 1. Find largest stationary portion of the signal.

Require N_s (odd number), α_L.
Ensure C, L, R.
  Partition the vibration signal into N_s blocks;
  $C \leftarrow \lceil N_s/2 \rceil$; ▷ central block
  $R \leftarrow 0$;
  for $L=1$ to $\lfloor N_s/2 \rfloor$ **do**
    Apply Levene's test, with significance level α_L, for blocks $C-L$ to $C+R$;
    if H_0 is accepted **then**
      $L \leftarrow L+1$;
    else
      $L \leftarrow L-1$;
      Break;
    end if
  end for
  for $R=1$ to $\lfloor N_s/2 \rfloor$ **do**
    Apply Levene's test for blocks $C-L$ to $C+R$;
    if H_0 is accepted **then**
      $R \leftarrow R+1$;
    else
      $R \leftarrow R-1$;
      Break;
    end if
  end for

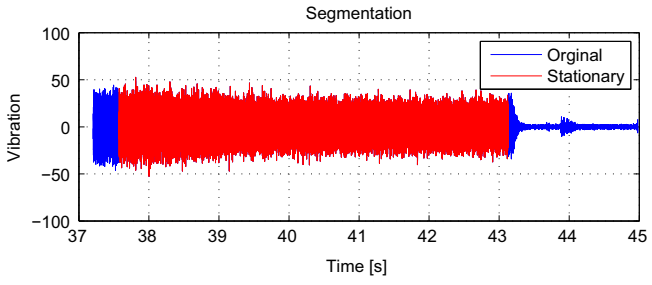

Fig. 4. Detection of the stationary portion of an individual run of the vibration signal: individual run (blue), stationary portion (red). (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

3.2. Multivariate statistical analysis

3.2.1. The T^2 distance method

The T^2 distance method is well established in the industry for multivariate analysis. Assume the availability of a data training set, consisting of n observations of m process variables, recorded in homogenous and nominal (*i.e.*, in absence of faults) operating conditions and stored in a matrix $X \in \mathbb{R}^{n\times m}$. The statistical distance T^2 of a generic measurement vector x from the training set is defined as $T^2 = x^T S^{-1} x$, where $S \in \mathbb{R}^{m\times m}$ is the (sampled) covariance matrix associated to X. In the ideal case, and under the assumption that variables x_i are Gaussian variables with zero mean and unitary variance, T^2 is characterized by a χ^2_m distribution, where m is indicated to denote the degrees of freedom of χ^2. Given a significance level α, we define $T^2_\alpha = \chi^2_m(\alpha)$ verifying $\mathbb{P}(T^2 \le T^2_\alpha) = \alpha$, which in turn denotes the ellipsoidal region $x^T S^{-1} x \le T^2_\alpha$.

The system is considered *out of control* (due to malfunctions or modified operating conditions) if $f_{out} > (1-\alpha)$, where f_{out} denotes the frequency of tested samples whose distances are outside the confidence region defined by T^2_α. The main steps of the method are listed below:

1. *Pre-processing*: Detect and remove any outliers. Then, normalize all variables to have zero mean and unitary variance: $x_{in} = (x_i - \mu_i)/\sigma_i$, where μ_i is the mean value and σ_i the standard deviation of x_i.
2. *Covariance calculation*: Compute the (sampled) covariance matrix S.
3. T^2 *distance calculation*: For any new set of measurements, compute the corresponding T^2 distances and test if $T^2 \le T^2_\alpha$ for at least $\alpha\%$ of the samples. In the negative case, the system is considered out of control.

In more realistic assumptions, when only a sampled covariance matrix is available, the most appropriate distribution to be considered is the F distribution and we define

$$T^2_\alpha = \frac{m(n-1)(n+1)}{n(n-m)} F_\alpha(m, n-m),$$

where $F_\alpha(m, n-m)$ is the $\alpha\%$ of the distribution F with m and $n-m$ degrees of freedom. The presented formulation of the T^2 distance can be extremely effective in the detection of anomalous operating conditions, although it cannot be used to isolate and identify the specific faults that caused such conditions.

3.2.2. Feature selection and training set

We consider a different training set for each vibration measurement and each run. For the definition of the statistical T^2 distance, the n observations of the process variables are obtained by dividing each signal into n subsequences of N_p samples, where N_p should be sufficiently large to eliminate noise-related effects. Also, the $m=6$ process variables consist of the following set of significant statistical features (Lebold, McClintic, Campbell, Byington, & Maynard, 2000) of the considered vibration signal.

- The mean value μ.
- The standard deviation σ.
- The skewness index $s = E[(x-\mu)^3]/\sigma^3$, which measures the asymmetry of the data with respect to their sampled mean value. If the distribution is shifted to the left with respect to the sampled mean value, the index is negative, and positive otherwise.
- The kurtosis index $k = E[(x-\mu)^4]/\sigma^4$, which measures the extent to which a distribution is subject to outliers. The kurtosis value corresponding to a normal distribution equals 3. Higher values denote a higher sensitivity to outliers.

- The crest factor, defined as $C = |x|_{peak}/x_{RMS}$, *i.e.* as the ratio of the peak value $|x|_{peak} = \max(|x|)$ to the average value x_{RMS} of the signal, calculated with the Root Mean Square (RMS).

Other features can be added at the cost of an increased computational effort, but this is not always effective, due to correlation effects. Principal Components Analysis (Jolliffe, 2002; Malhi & Gao, 2004) can be used to reveal the most significant features. More specifically, the eigenvalues of the covariance matrix S provide a measure of the significance of the corresponding features. A feature with an associated eigenvalue much smaller than the others should be removed because it does not carry new information.

3.2.3. Fault detection algorithm

The complete algorithm for batch fault detection on the given application consists of an off-line and an on-line phase, as explained below:

Algorithm 2. Fault detection algorithm: off-line phase.

Require α, N_p.
Ensure T^2_α.
  Collect data in nominal operating conditions.
  Segment the runs using the current absorption signal (see Section 3.1.1).
  for each vibration signal and each run **do**
    Remove any outliers from the data.
    Identify the stationary portions of the considered run (see Section 3.1.2).
    Split the signal into n sub-sequences of N_p points each.
    Compute the aggregate features for each sub-sequence.
    Build the (normalized) training set X.
    Compute and store the covariance matrix S.
    Compute the distance threshold T^2_α.
  end for

Algorithm 3. Fault detection algorithm: on-line phase.

Require T^2_α, N_p, μ, σ.
  Collect one batch of data for testing.
  Segment the runs using the current absorption signal.
  for each vibration signal and each run **do**
    Remove any outliers from the data.
    Identify the stationary portions of the considered run.
    Split the signal into n sub-sequences of N_p points each.
    for each subsequence **do**
      Normalize the features using μ and σ of the training set.
      Compute the T^2 distance.
      Compute the percentage of subsequences with $T^2 > T^2_\alpha$. If such percentage is greater than $(1-\alpha)\%$, declare a fault.
    end for
  end for

4. Fault isolation through envelope analysis

4.1. Envelope analysis

The previously described procedure is very effective in detecting an anomaly in the system, but less so in identifying the actual origin of the fault. Importantly, the specific types of fault can be identified if the fault frequency is detected.

In fact, various types of bearing defects (*e.g.*, ball or cage damages or inner/outer race defects) affect the acceleration signal in a similar way, *i.e.*, by generating an additive pulse train. Such faults manifest themselves at characteristic frequencies which can be very precisely related to the rotation frequency and are divulged by the manufacturer (see, *e.g.*, Randall, 2011). In view of this, revealing the fault characteristic frequency is equivalent to identifying the nature of the fault.

Estimating such frequency is far from trivial, since the amplitude of the fault additive pulse train can be small. A suitable technique, employed in the following, is called envelope analysis (Randall & Antoni, 2011). It allows to discriminate low frequency components, which would be completely masked with a simple Fourier analysis. The method consists of the following steps:

1. Band-pass filter the signal.
2. Compute the envelope of the filtered signal.
3. Compute the FFT of the obtained signal.

The resulting FFT will contain only the contributions associated to the modulating signal, *i.e.* to the faults of the bearings.

A crucial component of envelope analysis is the band-pass filter (*i.e.*, step 1) that must be designed to emphasize the contribution of the fault to facilitate its detection. Such filtering must eliminate both low-frequency components, associated, *e.g.*, with the shaft rotation and high frequency noise. The spectral kurtosis can be used to solve the problem of estimating an optimum detection filter directly from the data, without any *a priori* knowledge of the signal to be recovered (Randall & Antoni, 2006; Sawalhi & Randall, 2004). As recalled previously, the kurtosis is a statistical quantity (*i.e.*, the 4th moment of a distribution), which basically measures the distance of the underlying distribution from Gaussianity. Similarly, the spectral kurtosis (SK) is a spectral statistic, which ideally takes zero values at those frequencies where only Gaussian noise is present and large values where the transients occur. It can be shown, see Ge et al. (2010), that such behavior is closely related to the local signal-to-noise ratio. Therefore, the maximization of the SK is expected to maximize the signal-to-noise ratio as well.

The computation of the SK is based on the short-term Fourier transform (STFT) of the signal and is carried out as follows:

1. Divide the signal into M non-overlapping subsequences (each containing N_w samples).
2. Compute the N_w-point FFT $X_i(m)$ for each subsequence, $i = 1, \ldots, M$.
3. Compute the SK for each frequency beam m using the following unbiased estimator:

$$\mathrm{SK}_{N_w}(m) = \frac{M}{M-1}\left[\frac{(M+1)\sum_{i=1}^{M}|X_i(m)|^4}{\left(\sum_{i=1}^{M}|X_i(m)|^2\right)^2} - 2\right]$$

As also discussed in Ge et al. (2010), the band-pass filter should be designed so as to maximize the spectral kurtosis of the envelope of the filtered signal. It can be shown that this problem is strictly equivalent to finding the frequency $\overline{f}$ and the window length $\overline{N}_w$, which minimizes the STFT-based spectral kurtosis over all possible choices. The 3D map representing the spectral kurtosis over the frequency and the window length is called kurtogram. Once the frequency and the window length which maximize the kurtogram have been found, the central frequency of the band-pass filter is defined as $f_c = \overline{f}$ and its bandwidth as $B_w = 2f_s/N_w$, where f_s is the sampling frequency.

The Fourier transform of the envelope of a (filtered) faulty signal contains a periodic peak pattern, the distance between consecutive peaks representing the base frequency of the fault. An algorithm is here proposed to robustly extract the fault frequency.

Table 1
Values of the user-defined parameters.

Parameter	Step	Value
β	Individual run identification (Section 3.1.1)	0.5
N_s	Algorithm 1	21
α_L	Algorithm 1	0.1
N_p	Algorithms 2 and 3	1500
α	Algorithms 2 and 3	0.99
F_{min}	Algorithm 4	5 Hz
F_{max}	Algorithm 4	500 Hz
$\overline{c}_F$	Algorithm 4	0.05

First, a threshold for the peak detection is established through an iterative procedure. Then, the distances between the peaks are processed to estimate the fault frequency, employing statistical reasoning. In detail, the Procedure sketched in Algorithm 4 must be carried out on the spectrum.

Algorithm 4. Fault frequency estimation algorithm.

Require F_{min} and F_{max}, where $F_{min} < F_{max}$;
Ensure μ_F;
  Remove all the points below F_{min} and above F_{max};
  Compute the mean μ_M and the standard deviation σ_M of the spectrum;
  $k \leftarrow 2$; $valid \leftarrow false$;
  while $valid = false$ **do**
    $\gamma \leftarrow \mu_M + k\sigma_M$;
    Find the peaks above γ.
    if number of peaks < 3 **then**
      Break;
    end if
    Compute the distances in frequency between consecutive peaks;
    Compute the mean μ_F and the standard deviation σ_F of such distances;
    Remove all the distances outside the range $[\mu_F - \sigma_F, \mu_F + \sigma_F]$;
    Re-compute the mean μ_F and the standard deviation σ_F on the remaining distances;
    $c_F \leftarrow \sigma_F / \mu_F$;
    if $c_F < 0.05$ **then**
      $valid \leftarrow true$;
    else
      $k \leftarrow k + 0.5$;
    end if
  end while

The lower bound for the frequency (*i.e.*, F_{min}) is essentially needed to remove the DC peak, which is always present and does not carry any useful information. A reasonable value is $F_{min} = 5$ Hz, which removes also the sideband of the DC component. On the other hand, the upper bound F_{max} can be omitted. However, the peaks typically present a decreasing amplitude and thus above a certain frequency they are concealed by the background noise.

The outlier distances are identified as those exceeding the $\mu_F \pm \sigma_F$ thresholds. Although this may seem too strict a condition, it is justified upon observing that the peaks should be exactly separated by the fault frequency. Thus, such bound is suitable to remove errors caused by spurious components. Also, note that in view of the presence of noise and of numerical reasons, a difference is to be expected between the frequency distances assumed to be valid (*i.e.*, lying in the range $[\mu_F - \sigma_F, \mu_F + \sigma_F]$). To account for this a small tolerance level has been introduced, defined in terms of the coefficient of variation c_F, *i.e.*, the ratio between the standard deviation and the mean. The latter coefficient is a commonly used measure of the dispersion of a population with respect to its mean (Gastwirth et al., 2009). The threshold level for c_F has been tuned mostly based on empirical results, and it is here set to $\overline{c}_F = 5\%$.

If the data are considered invalid, the peak threshold is increased by incrementing the factor k. The algorithm stops if either less than 3 peaks are found, in which case the analysis does not provide any result, or if a coefficient of variation lower than 5% is found. In the latter case, the analysis provides a valid value of the fault frequency (μ_F), together with its standard deviation.

4.2. Varying shaft angular speed

The presented algorithm relies on the assumption that the shaft velocity is constant during the data processing, so that it is safe to assume that the fault frequency is constant during the run. Indeed, as discussed in Randall and Antoni (2011), the latter is (linearly) proportional to the angular velocity of the motor shaft. However, in practice, the shaft velocity can significantly deviate from this condition, causing the frequency peaks of the envelope spectrum to be irregularly spaced out, which prevents a correct detection of the fault.

This problem can be effectively tackled by representing the involved signal as a function of the motor shaft angle instead of the time. Precisely, the angle can be computed as a function of time by integrating the angular velocity: $\theta = h(t)$. Then, signal $y(t)$ can be expressed in the angle domain as

$$\overline{y}(\theta) = y(h^{-1}(\theta)).$$

The spectrum of the corresponding envelope displays regularly spaced peaks in the presence of a fault even in case the shaft speed is not constant. Therefore, Algorithm 4 can be directly applied on the spectrum of the envelope of $\overline{y}(\theta)$, after the appropriate filtering action. If a fault is detected, the associated "angular periodicity" is equal to $\theta = 1/\mu_F$, where μ_F is the average distance between consecutive peaks, and the average fault frequency can be estimated as

$$f_{fault} = \frac{\overline{\omega}}{\theta},$$

where $\overline{\omega}$ is the average angular frequency in the considered signal portion.

5. Analysis of experimental data

In this section we test the ability of the algorithm to detect anomalous operating conditions. To this aim, artificial faulty vibration signals have been synthesized and added to real measurements.

Bearing defects typically manifest themselves as quasi-cyclostationary phenomena, which can be quite closely approximated as periodically repeating filtered impulses (McInerny & Dai, 2003; Randall, 2011; Randall & Antoni, 2011; Sawalhi & Randall, 2008). Accordingly, the faulty signals have been simulated by corrupting a vibration measurement with an additive train of pulses, the characteristic frequencies of which are determined by the geometry of the bearing and the rotational speed of the shaft (Randall, 2011). Also, as discussed in Ge et al. (2010), the signal modeling the fault should be filtered in order to simulate the filtering effect of real environ-ments (bearing connections, mechanical joints, machine shell). In the following examples, the pulse trains have been filtered by a second order system with complex-conjugate poles with natural frequency $f_p = 2.5$ kHz.

The values of the tuning parameters, adopted in the present example, are given in Table 1.

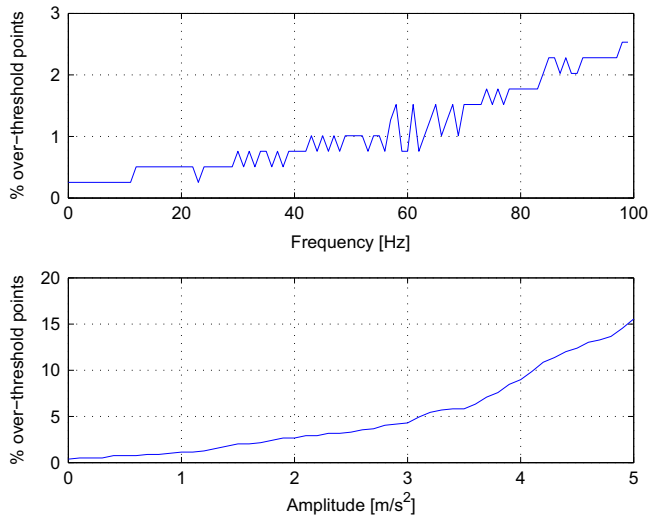

Fig. 5. Algorithm sensitivity to the fault frequency (top) and amplitude (bottom).

Table 2
Percentage of the segments of the vibration signal outside the confidence region in nominal conditions.

Data set	Vibration signals						
	1	2	3	4	5	6	7
A	0.7	0.7	0.0	0.0	0.0	5.1	0.0
B	0.0	0.7	1.5	0.0	0.7	0.7	1.5

Table 3
Percentage of the segments of the vibration signal outside the confidence region in faulty conditions.

Data set	Vibration signals						
	1	2	3	4	5	6	7
A	5.8	29.7	2.9	29.7	29.7	11.6	29.7
B	0.7	31.2	7.2	31.2	9.4	29.7	31.2

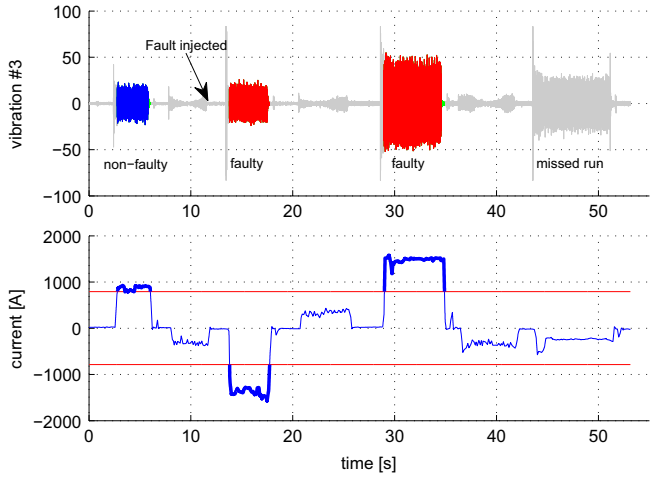

Fig. 6. Complete processing of a data-set with the proposed algorithm: vibration signal (top) and current signal (bottom).

5.1. Sensitivity of multivariate statistical analysis to the fault amplitude and frequency

First of all, it is important to establish the sensitivity of the preliminary fault detection phase to the amplitude and frequency of the bearing fault. Fig. 5 (top) reports the percentage of signal segments with a T^2 distance exceeding the threshold defined by a significance level $\alpha = 99\%$ as a function of the fault frequency (the corresponding fault amplitude is set to 1 m/s^2). Apparently, the higher the frequency, the more points exceed the bound. However, the fault can be detected even for very low frequencies. The low frequency threshold is also motivated by the size of the sub-sequences. In the examined case, each sub-sequence contains 1000 samples, which correspond to about 42 ms at a sampling frequency of 24 kHz. Therefore, in order to have at least one pulse in each sub-sequence, the train frequency must be greater than 24 Hz. Below such frequency the algorithm cannot identify any fault. The sensitiv-ity of the algorithm to the amplitude of the faults has been tested as well, setting the frequency of the pulse train to 80 Hz (see Fig. 5, bottom). Apparently, the fault can be detected even for very small amplitudes, demonstrating the effectiveness of the fault detection algorithm even in the early stage of the defect progress.

5.2. Step I: Multivariate statistical analysis

A total of 16 sets of real-data measurements, each pertaining to an entire bar processing, have been collected over a large period of time. Bars of different length and steel quality are considered in the data set. Three of these data sets, representing typical working conditions, have been chosen to build the training set. In the following, for simplicity, we will make reference only to the first two data sets not belonging to the training set (denoted by *A* and *B*). However, the same conclusions can be drawn for the remaining data sets. The first experiment aims to show the effect of a fault on the statistical distance T^2. The entries in Tables 2 and 3 represent the percentage of signal segments (for each vibration signal) whose T^2 distance is greater than the threshold. In particular, Table 2 has been obtained in "safe" operating conditions, *i.e.* using the raw vibration signal with-out any additive faults. By contrast, Table 3 has been obtained in faulty conditions, with a fault amplitude of barely 1 m/s^2, and a fixed frequency of 80 Hz. The difference is clearly visible, leading to easy fault detection by a simple threshold-checking criterion. Note that the increment of the percentage values caused by the presence of the fault is not uniform among the signals: this is due to the difference in amplitude of each signal, which is caused by the different positions of the bearings in the reducer.

In the second experiment we tested the complete batch algorithm, from the data segmentation based on the current signal to the actual fault detection. Fig. 6 illustrates the segmentation of the current signals and the resulting segmented vibration signal portions (the colored parts), with reference to signal 3 of data set *A*. As previously discussed, the fourth run is missed. The faulty signal (with amplitude 5 m/s^2) is injected only in between the 1st and 2nd runs, thus leaving the first one in safe conditions. The algorithm, besides the proper segmentation, correctly identifies the faults, whenever they are present.

To avoid possible false alarms as well as missed faults, the robustness of the method could be further improved by consider-ing repetitive patterns in the fault detection process. More precisely, one can adopt the simple rule to declare a fault only if the signal appears to be faulty for a certain number of consecutive runs, following the same idea of the Western Electric Rules (Montgomery, 2009).

To better appreciate the test discussed above, we also report the T^2 distances for each run in Fig. 7. The upper-most figure shows the distances for each segment of the first run, during

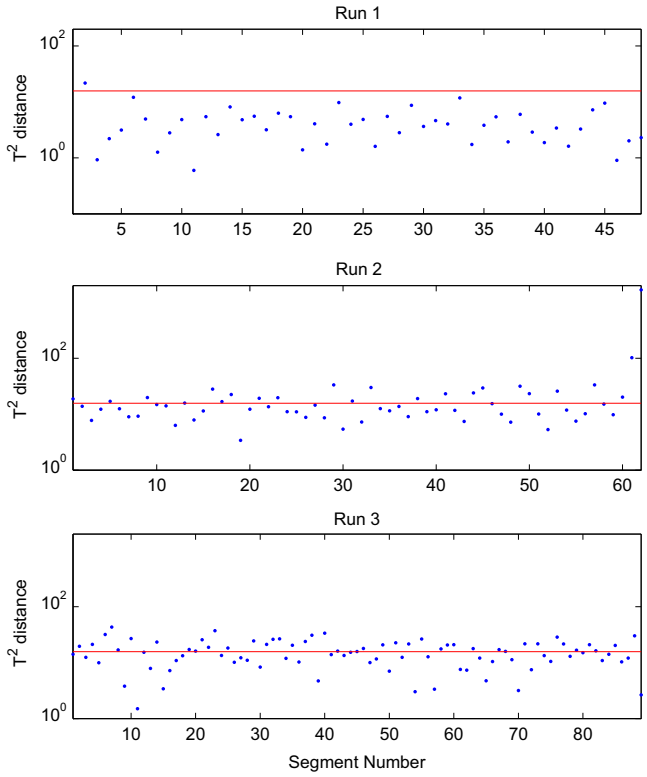

Fig. 7. T^2 distance evaluation for the three runs of the data-set considered in the example.

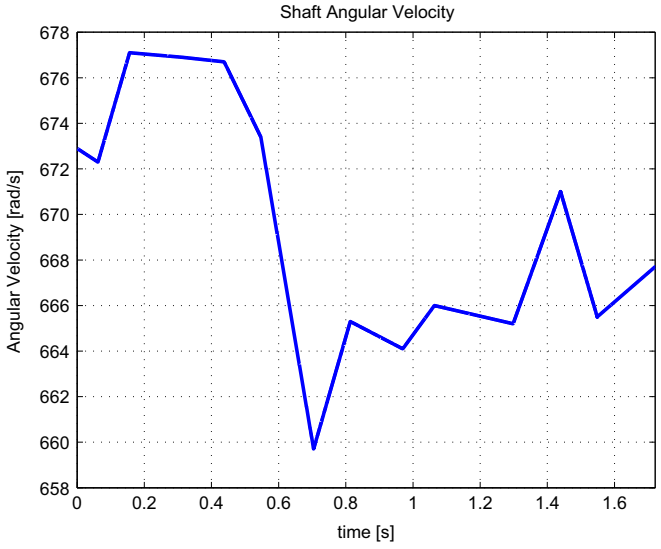

Fig. 8. Shaft angular velocity.

which the faulty signal was not yet injected. In accordance with the results of Tables 2 and 3, the majority of the distances are below the threshold (in red), correctly resulting in no faults being detected. In contrast, during runs 2 and 3 the faulty pulse train was added leading to a much bigger number of segments with T^2 distance greater than the allowed threshold.

5.3. Step II: Envelope analysis

The envelope analysis method has been tested on signal 1 during run 1. The bearing fault has been modeled as a filtered pulse train of amplitude 7 m/s^2 occurring two times for each shaft

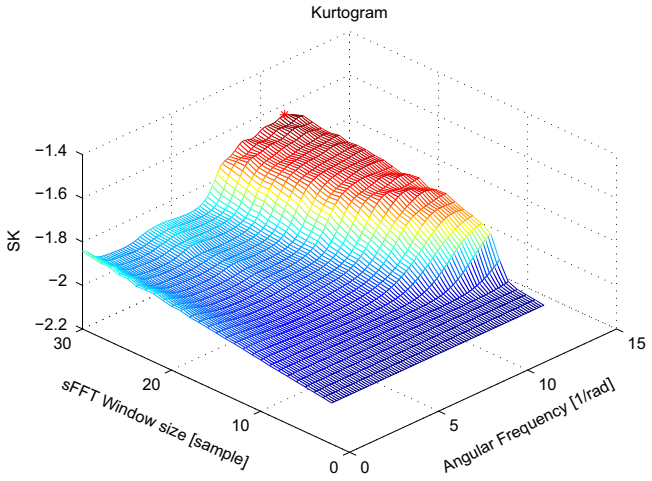

Fig. 9. Kurtogram.

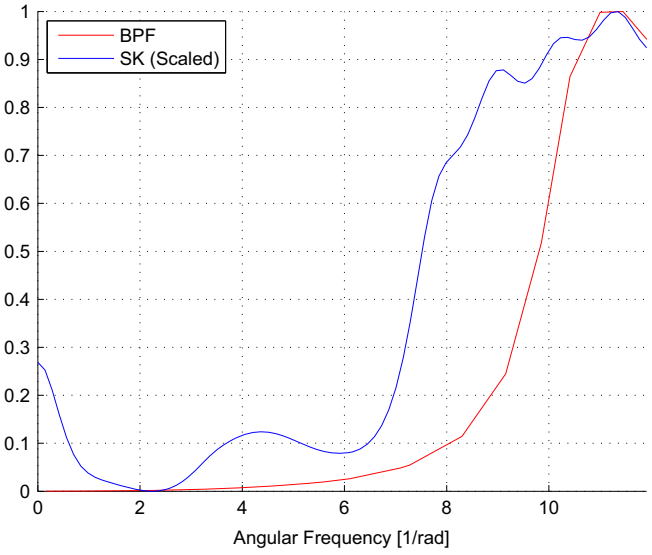

Fig. 10. Band pass filter (red) and spectral kurtosis (blue). (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

rotation, *i.e.*, $\bar{\theta} = \pi$. In view of the fact that the shaft velocity is time-varying (as witnessed by Fig. 8), we resort to the time-to-angle reformulation of the vibration signal before applying Algorithm 4.

Fig. 9 shows the kurtogram obtained as described previously. The maximum value of the spectral kurtosis corresponds to the angular frequency 11.35 rad^{-1} and N_w=30 samples. Therefore, the resulting band pass filter has a central periodicity f_c=11.35 rad^{-1} and a bandwidth $B_w = 2f_s/N_w = 2.3914$ rad^{-1} (in the angular domain f_s=35.8717 rad^{-1}).

Fig. 10 shows the frequency response of the resulting band pass filter together with the spectral kurtosis corresponding to the optimum window length, which has been properly scaled to fit the figure.

Algorithm 4 has been applied to the Fourier Transform of the envelope of the filtered signal, the outcome being illustrated in Fig. 11. The algorithm identifies a fault angular periodicity of 0.318 rad^{-1} with a standard deviation of 0.0006 rad^{-1}, providing satisfactory result (recall that the exact periodicity is $1/\pi = 0.3183$ rad^{-1}).

To show the importance of performing the time-to-angle reformulation of the vibration signal in case of varying shaft

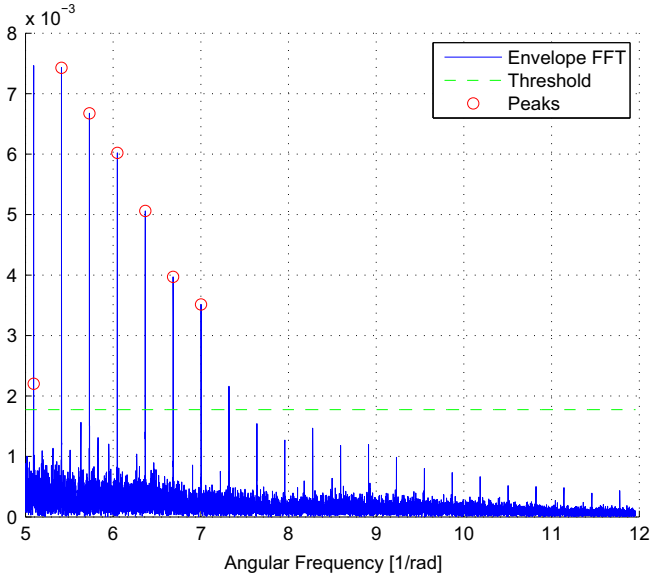

Fig. 11. Envelope analysis.

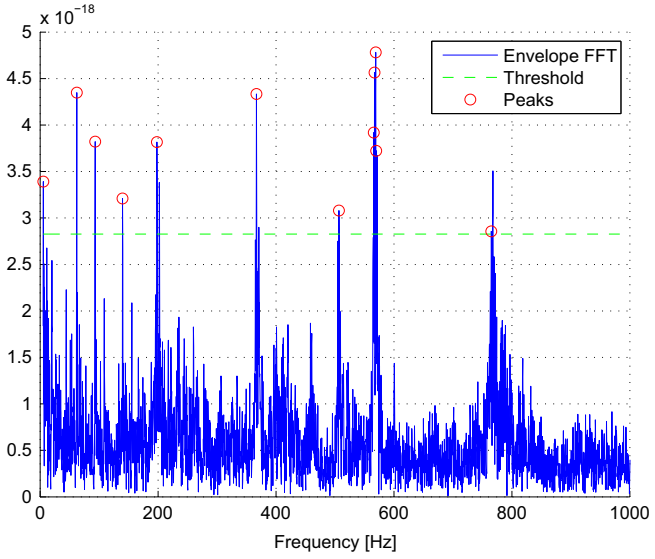

Fig. 12. Envelope analysis obtained by analyzing the original time-domain signal.

speed, we have reported in Fig. 12 the behavior of Algorithm 4 when directly applied to the time-domain signal. Apparently, even the small variability of the shaft angular velocity observed in the data dramatically affects the outcome of the SK analysis. This is witnessed by the fact that, as a result of a wrong frequency detection, the band-pass filter results badly tuned, and the amplitude of the spectrum is extremely small (see Fig. 12).

Finally, the envelope analysis has been tested on a non-faulty signal to evaluate its ability to avoid false alarms. The results are depicted in Fig. 13. In the absence of the fault the spectrum of the envelope does not show any periodic behavior. Therefore, the algorithm is not able to find a group of peaks whose distances are constant, and correctly reports a no-fault condition.

6. Conclusions

In this work we have proposed a two-step scheme, relying on two complementary data-driven techniques, for fault detection and isolation for a drive reducer in a hot steel rolling mill.

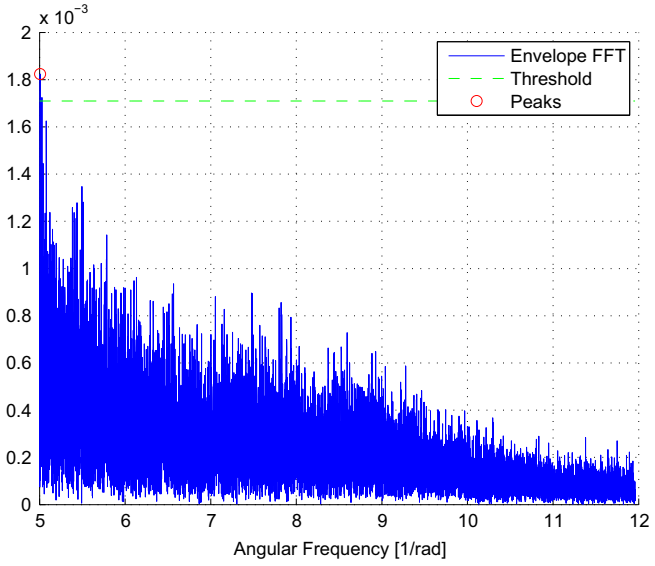

Fig. 13. Envelope analysis in no fault conditions.

A preliminary fault detection phase is carried out based on a simple time-domain analysis algorithm. This method, based on multivariate statistical analysis and on Hotelling's T^2 distance, appears to be very effective for the detection of faults. In contrast to other methods (*e.g.*, frequency-based), it is computationally lightweight and leads to a very simple limit checking as fault detection rule.

Secondly, a more complex and computationally intensive method, based on envelope frequency-domain analysis, is then used to confirm the detection of the fault and provide more detailed information on its frequency characteristics.

In the paper, we provide automatic procedures for the practical application of the proposed techniques. To test the effectiveness of the proposed scheme, classical bearing defect models are employed, in combination with real-data measurements. Future work will focus on the analysis of fault detectability from accelerometers not directly associated with the faulty bearing.

References

Baydar, N., Chen, Q., Ball, A., & Kruger, U. (2001). Detection of incipient tooth defect in helical gears using multivariate statistics. *Mechanical Systems and Signal Processing*, *15*(2), 303–321.

Ericsson, S., Grip, N., Johansson, E., Persson, L.-E., Sjöberg, R., & Strömberg, J.-O. (2005). Towards automatic detection of local bearing defects in rotating machines. *Mechanical Systems and Signal Processing*, *19*(3), 509–535.

Gastwirth, J. L., Gel, Y. R., & Miao, W. (2009). The impact of Levene's test of equality of variances on statistical theory and practice. *Statistical Science*, *24*(3), 343–360.

Ge, Z., Kruger, U., Lamont, L., Xie, L., & Song, Z. (2010). Fault detection in non-Gaussian vibration systems using dynamic statistical-based approaches. *Mechanical Systems and Signal Processing*, *24*, 2972–2984.

Jolliffe, I. (2002). *Principal component analysis*. New York, USA: Springer Verlag.

Lebold, M., McClintic, K., Campbell, R., Byington, C., & Maynard, K. (2000). Review of vibration analysis methods for gearbox diagnostics and prognostics. In *Proceedings of the 54th meeting of the society for machinery failure prevention technology*, Virginia Beach (VA), USA (pp. 623–634), May 1–4.

Malhi, A., & Gao, R. X. (2004). PCA-based feature selection scheme for machine defect classification. *IEEE Transactions on Instrumentation and Measurement*, *53*(6), 1517–1525.

McInerny, S., & Dai, Y. (2003). Basic vibration signal processing for bearing fault detection. *IEEE Transactions on Education*, *46*(1), 149–156.

Montgomery, D. C. (2009). *Introduction to statistical quality control*. (6th ed.). New York (NY), USA: John Wiley & Sons.

Randall, R. B. (2011). *Vibration-based condition monitoring: Industrial, aerospace and automotive applications*. New York (NY), USA: John Wiley & Sons.

Randall, R. B., & Antoni, J. (2006). The spectral kurtosis: Application to the vibratory surveillance and diagnosis of rotating machines. *Mechanical Systems and Signal Processing*, *20*, 308–311.

Randall, R. B., & Antoni, J. (2011). Rolling element bearing diagnostics—A tutorial. *Mechanical Systems and Signal Processing*, *25*, 485–520.

Sawalhi, N., & Randall, R. B. (2004). The application of spectral kurtosis to bearing diagnostics. In *Proceedings of the acoustics*, Gold coast, Australia (pp. 393–398), November 3–5.
Sawalhi, N., & Randall, R. B. (2008). Simulating gear and bearing interactions in the presence of faults. Part I. The combined gear bearing dynamic model and the simulation of localized bearing faults. *Mechanical Systems and Signal Processing*, *22*, 1924–1951.
Zhang, B., Sconyers, C., Byington, C. S., Orchard, R. P. M. E., & Vachtsevanos, G. J. (2011). A probabilistic fault detection approach: Application to bearing fault detection. *IEEE Transactions on Industrial Electronics*, *58*(5), 2011–2018.