Bayesian inference for outlier detection in vibration spectra with small learning dataset

Aurélien Hazan, Michel M. Verleysen, Marie Cottrell, Jérôme Lacaille

Bayesian inference for outlier detection in vibration spectra with small learning dataset

Aurélien Hazan[a], Michel Verleysen[a,b], Marie Cottrell[a], Jérôme Lacaille[c]

[a] *SAMM - Université Paris 1*
[b] *DICE - Université Catholique de Louvain*
[c] *SNECMA - Division Intégration, Villaroche*

Abstract

The issue of detecting abnormal vibrations is addressed in this article, when little is known both on the mechanical behavior of the system, and on the characteristic patterns of potential faults.

With data from a bearing test rig and from an aircraft engine, we show that when only a small learning set is available, Bayesian inference has several advantages in order to compute a model of healthy vibrations, and thus ensure fault detection.

To do so, we compute the wavelet transform of many log-periodograms, and show that their probability density can be easily modelled. This allows us to compute a likelihood index when a new log-periodogram is presented, thanks to marginal likelihood approximation. A by-product of this computation is the ability to generate random log-periodograms according to the learning dataset probability density.

Finally, we first detect the degradation of a bearing on a test rig; then we generate random samples of aircraft engine log-periodograms.

Keywords: fault detection, vibration, rotating machine, bearing, aircraft engine, bayesian inference, periodogram, wavelet, probabilistic model

---

1. Introduction

We tackle the issue of prognosis and health monitoring for rotating machine. Our goal application is vibration monitoring in aircraft engines, but

---

Email address: aurelien.hazan@univ-paris1.fr (Aurélien Hazan)

simpler test cases are also dealt with, for example bearing test rig. The following hypotheses are made:

- small learning set: no more than a dozen of short time-series are available, as a reflect of industrial constraints. More specifically, faulty data are scarce, if any.
- model-free: no specific mechanical model of the system, nor model of faults that might occur are used in the following.
- constant target rotation speed: the rotating machines studied in this article have a fixed speed. On short time intervals, the signal will be supposed stationary, so that periodograms are meaningful.
- nonparametric estimation: no specific functional form is assumed concerning periodograms, which will be decomposed in a wavelet basis.

In the spirit of many works in Novelty Detection (Markou and Singh, 2003a) where information on faulty data is limited, our aim is: first to come up with a nonparametric model of a healthy signal, using a small learning set; secondly to compare any new signal with this model; lastly to detect unusual behavior.

To build this model of healthy vibration signals, we consider the wavelet transform of periodograms, which offers enough freedom in the perspective of function approximation. The wavelet transform of periodograms has been the subject of many works in statistical estimation and signal processing, as a means to denoise periodograms (see 2). In this litterature, a bayesian point of view is adopted, such that all learnt coefficients (and therefore the full periodogram) have a probabilistic description. This feature will be useful for detection, as we will see in the following: section 2 links our work to related articles in various fields, sec. 3 presents the main algorithms we use, which results are summarized in sec. 4. Sec. 5 concludes this article and discusses its perspectives.

2. Related work

Vibratory Health monitoring applied to rotating machines often focusses on specific faults, such as rotor/stator contact (Peng et al., 2005), rotor unbalance, blade defects (Kharyton, 2009), bearing (Orsagh et al., 2003) and gearings defects (Wang et al., 2001).

However, unexpected problems can occur, whose fault patterns are unknown. Such preoccupations are germane to those developped in the area of novelty detection, where the importance of data not seen during the learning phase is stressed. Facing this problem, the best solution found by many authors was to build a model of normality, for example with neural networks such as Self-Organising Maps (Ypma and Duin, 1997; Tarassenko et al., 1999; Markou and Singh, 2003b). More probabilistic approaches exist (Markou and Singh, 2003a), where a probabilistic model of normality is rather built. This articles will adopt this point of view.

For condition monitoring, signals may be studied in various domains: time domain, the Fourier basis via STFT, the wavelet domain (Peng and Chu, 2004), or other time-frequency distributions such as Wigner-Ville (Antoni et al., 2004).

More specifically, we propose to build a probabilistic model of normality of log-periodograms in the wavelet domain. This is justified by the fact that such models were developped in statistical time series analysis and signal processing, for spectrum denoising purposes (Moulin (1994), Percival and Walden (2000, 10.6), Vidakovic (1999, 9.3), Pensky et al. (2007)), via wavelet thresholding or shrinking. The motivation of researchers in this area concern mainly the statistical properties of estimators (such fixed or variable bandwith smoothing), which we will not discuss here. However we propose to take advantage of the model of normality that is provided by their analysis.

In a similar spirit, a Bayesian approach to normality modelling in jet engine health monitoring has been developped (Clifton, 2006; Clifton et al., 2008; Clifton and Tarassenko, 2009), but so far only a restricted number of shaft order amplitudes are considered, and not the whole spectrum. These works are very interesting because they use Extreme Value Theory to model the maxima of order amplitudes, and set thresholds that are more robust.

Before continuing, we now briefly discuss two objections that the reader might formulate:

- why not working directly with the wavelet transform of the time-domain signals, without computing its periodogram ? Because we would have no simple probabilistic model of the coefficients.

- why not working directly with the probabilistic models of the log-periodograms ? Indeed, as we will see in sec. 3, such a probabilistic model is available, under stationarity assumptions. However, the noise distribution is not standard (see 3.2), which complicates subsequent computations.

In sec. 3, we discuss the various algorithms that aim at inferring a normality model, then exploiting it to detect faults.

3. Algorithms: Bayesian Detection and random spectra generation in a wavelet basis

Basic familiarities with the wavelet transform, and its discrete implementation is assumed in this section. Theorical foundations, principles of fast computation, as well as practical illustrations may be found in (Mallat, 2009).

3.1. Probabilistic model of the wavelet transform of a periodogram

What follows is standard material, available from (Moulin, 1994). We adopt the notations of (Pensky et al., 2007).

Let $I(\omega_j)$ be the periodogram at Fourier frequency $\omega_j = \frac{2\pi j}{T}$ associated with vibration signals $X_0, \ldots, X_{T-1}$:

$$I(\omega_j) = \frac{1}{2\pi T}|\sum_{t=0}^{T-1} X_t e^{-i\omega_j t}|^2 \tag{1}$$

$I(\omega_j)$ is an estimator of power spectrum density (PSD) which probability density function can be approximated under mild stationarity assumptions (Brillinger, 1981), as a function of the true PSD, $f(\omega_l)$:

$$I(\omega_l) \overset{iid}{\approx} \frac{1}{2} f(\omega_l)\chi_2^2 \tag{2}$$

where ω_l is distinct from the extremities. Taking the log, a regression formula can be proposed :

$$z_l = \ln f(\omega_l) + \varepsilon_l \tag{3}$$

where $z_l = \ln I(\omega_l) + \gamma$, and γ is Euler's constant. Let μ the density of ε_l. It can be shown that:

$$\mu(x) = \gamma^* \exp(x - \gamma^* e^x) \tag{4}$$

$$E[\varepsilon_l] = 0 \tag{5}$$

$$V[\varepsilon_l] = \frac{\pi^2}{6} \tag{6}$$

where $\gamma^* = e^{-\gamma}$.

Taking the discrete wavelet transform of eq. (3):

$$d = \theta + \delta \tag{7}$$

where:

$$\begin{aligned} d &= W[z_1, \dots, z_T] && (8) \\ \theta &= W[\ln f(\omega_1), \dots, \ln f(\omega_T)] && (9) \\ \delta &= W[\varepsilon_1, \dots, \varepsilon_T] && (10) \end{aligned}$$

and W is an orthogonal matrix given by the discrete wavelet transform. d, θ and δ may also be indexed by (j, k), where j is the scale and k the position.

By Central Limit Theorem arguments, the density of coefficients of vector δ can be approximated by a normal law, except for small scales where a correction must be applied.

3.2. Bayesian inference of a logperiodogram, in the wavelet domain

Assuming the model of sec. 3.1, what can be learnt from measurements on the distribution of wavelet coefficients θ ?

Here we assume a Bayesian inference scheme, since it ensures that θ keeps a probabilistic description when data are available. Prior for the wavelet coefficient θ_{jk} of the following form may be found in the litterature (Pensky et al., 2007):

$$\theta_{jk} \rightsquigarrow \pi_j \delta(0) + (1 - \pi_j)\tau_j \xi(\tau_j \theta_{jk}) \tag{11}$$

where ξ is symmetric (such as a normal law $\mathcal{N}(0, 1)$), and π_j, τ_j are hyperparameters. They can be learnt independently, taking advantage or theoretical arguments. In this article, we simply use a centered normal prior with variance obeying the following model (Abramovich et al., 1998):

$$\sigma^2 = C 2^{-\alpha j} \tag{12}$$

where C and α are constats learnt from the data.

A posterior can be computed by the classical Bayes formula:

$$P\left(\theta_{jk} | d_{jk}^{(1)}, \dots, d_{jk}^{(n)}\right) \propto l_{\mathbf{d}_{jk}}(\theta_{jk}) Pr(\theta_{jk}) \tag{13}$$

Standard calculus shows that the posterior has the following form:

$$\forall(j,k), P\Big(\theta_{jk}|d^{(1)},\ldots,d^{(n)}\Big) \quad \propto \quad \exp\Big(-\frac{1}{2\sigma_0^2}\Big[\theta_{jk} - \frac{\widehat{d}_{jk}}{1+\frac{\sigma_1}{\sigma_2}}\Big]^2\Big) \tag{14}$$

where $\widehat{d}_{jk}$ is the mean wavelet coefficient of the sample periodograms, and $\sigma_0, \sigma_1, \sigma_2$ are standard deviations whose formula are given in Appendix A.

3.3. Random generation of log-periodograms

Once the distribution of the posterior in eq. (14) is computed thanks to log-periodogram samples, one can sample from this distribution. Computing the inverse wavelet transform, we get a random log-periodogram sample. Examples will be given in sec. 4.

Due to acquisition cost, such random samples can be of high interest to test detection algorithms. The classical bayesian fault detection procedure is highlighted in the following section.

3.4. Fault detection

So far we have chosen an estimation model (see eq .(7)), proposed a prior and a posterior (see eq. (14)). We can now compute the marginal likelihood (Bishop (2006, eq.(3.67-68)), Clifton et al. (2008, eq.(5))), which quantifies the likelihood of a new sample, given the training set

$$d \mapsto p(d|d^{(1)},\ldots,d^{(n)}) \quad = \quad \int p(d|\theta)p(\theta|d^{(1)},\ldots,d^{(n)})d\theta \tag{15}$$

The integral may be approximated by Monte-Carlo sampling ((Robert and Casella, 2010, 3.2). If $p(d|d^{(1)},\ldots,d^{(n)})$ is below a given threshold, a fault is suspected to occur.

4. Data and Results

4.1. Data: IMS bearing dataset

The IMS bearing dataset (Qiu et al., 2006) is a publicly available[1] set of vibration signals. Four bearing are installed on a shaft that rotates at a constant speed of 2000rpm. Progressive degradations are recorded from 8 accelerometers as the designed life time of the bearings is exceeded.

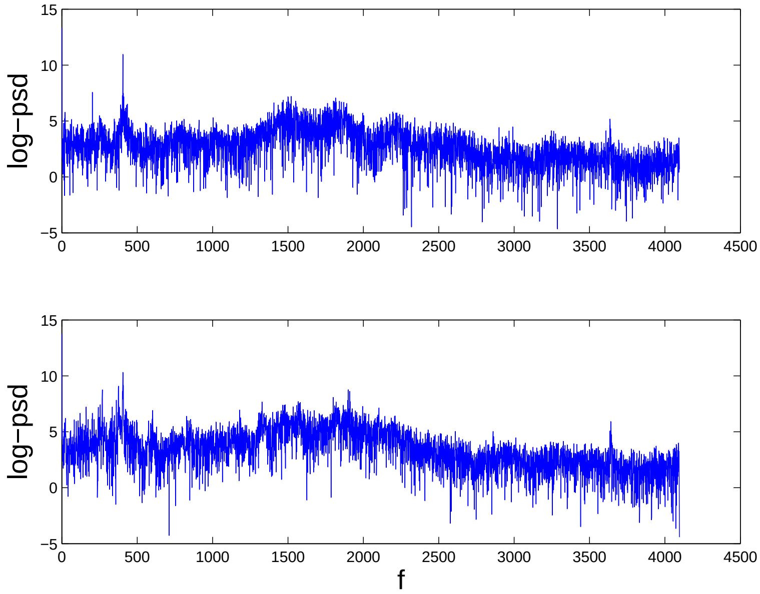


Figure 1: Log-periodogram of bearing vibrations (top) at the beginning of the test; (bottom) at the end of the test when bearing are damaged.

Log-periodograms are displayed by Fig. 1 at the beginning and at the end of the test, when the bearing is damaged.

Two datasets are built from the IMS recordings: one learning dataset, with snapshots taken at the start of the recording session, while all bearings are healthy. Then, a test dataset is designed with new recordings taken at the start, in the middle and at the end of the degradation test.

4.2. Data: Snecma turbofan engine

The recordings under study were provided by the Health Monitoring Department of SNECMA[2] and correspond to a dual-shaft turbofan mounted on a testbench, that undergoes a continuous acceleration during several minutes. They include raw vibration outputs of two embedded accelerometers, sampled at 51kHz. Samples are collected while low-pressure shaft speed is at 2000rpm. No failure dataset is available at the moment.

[1] `http://ti.arc.nasa.gov/tech/dash/pcoe/prognostic-data-repository/`

[2] `http://www.snecma.fr`

4.3. Implementation details

Discrete wavelet transform is computed thanks to Wavelab[3]. In the following, the Haar wavelet is used.

4.4. Results: IMS bearing dataset

In agreement with the processing steps depicted in sec. 3.2, hyperparameters learning is first evoked. The model eq.(12) is fitted thanks to empirical standard deviations of discrete wavelet coefficients. As outliers are suppressed, a good fit is obtained except at extremities, as illustrated by Fig. 2.

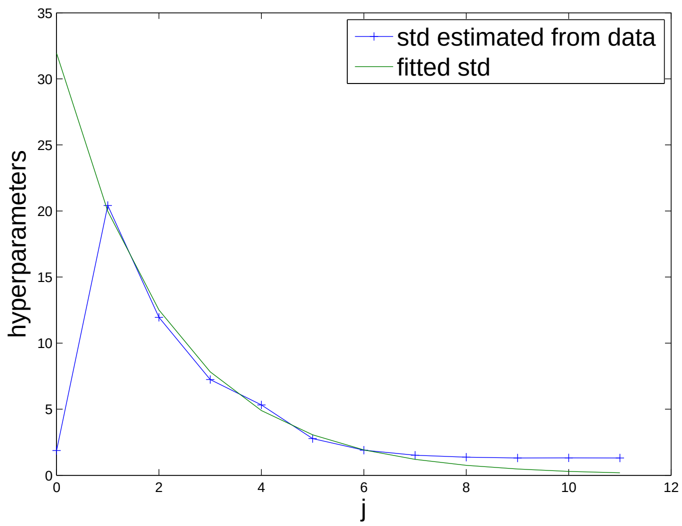


Figure 2: Hyperparameters fit.

Now we proceed to fitting posterior distributions as announced in sec. 3.2, thanks to the IMS learning subset, depicted in sec. 4.1. Fig. 3(top) represents the histograms of discrete wavelet coefficients of log-periodograms for a selection of (j, k) indices, with j low. Because of the low number of samples, which is part of the constraints mentionned in sec. 1, the shape of the histograms is hardly distinguishable. This is why the theoretical arguments evoked in sec. 3.1 in favour of normality, as well as the Bayesian updating are helpful to get the posteriors represented by Fig. 3(bottom).

[3] `www-stat.stanford.edu/~wavelab/`

We remark that, however incomplete the histograms, they do not display a large number of zeros, as could be expected from the classical prior eq. (11). This point will be discussed in sec. 5.

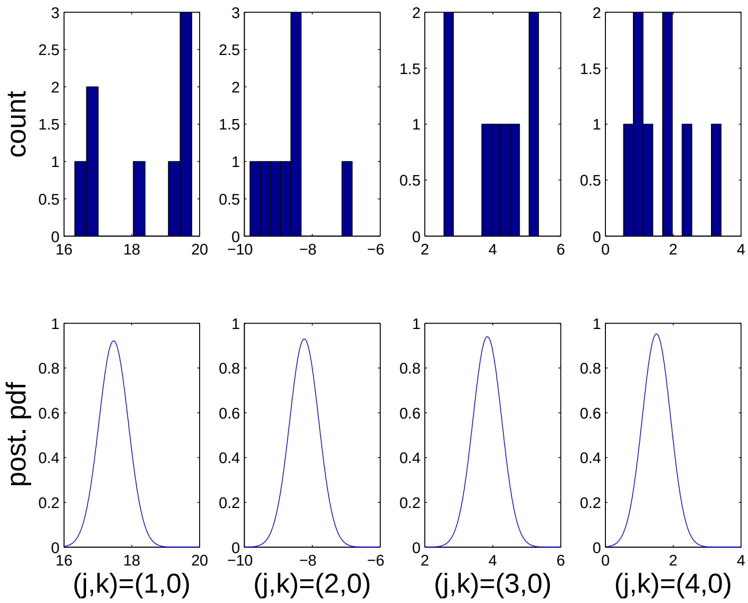


Figure 3: (top) Histogram of discrete wavelet coefficients of log-periodograms; (bottom) posterior densities, indexed by (j, k).

Sampling from this posterior as indicated in sec. 3.3, we obtain random log-periodograms as shown by Fig.4. The general shape is correct, thanks to coeffcients in low and intermediate scales. However, the spectrum is thicker, because of high scale coefficients, whose variance is higher, as was expected from the hyperparameters fitting step. This will be corrected in later works.

Finally we compute the marginal likelihood, which was proposed in sec. 3.4 to detect faults. To do so we use the test IMS subset described in sec. 4.1 which contains healthy recordings different from learning ones, and snapshots at partly then severely deteriorated bearings. We expect different values of marginal likelihood to be displayed as time passes.

Fig. 5 summarizes the results, that indeed reveal a clear distinction between the several stages of degradation. This validates our approach, according to which bayesian modelling of the wavelet transform of log-periodograms can detect degradation of the condition of bearings. Further experiments, and extensive comparison with state-of the art methods will be addressed in

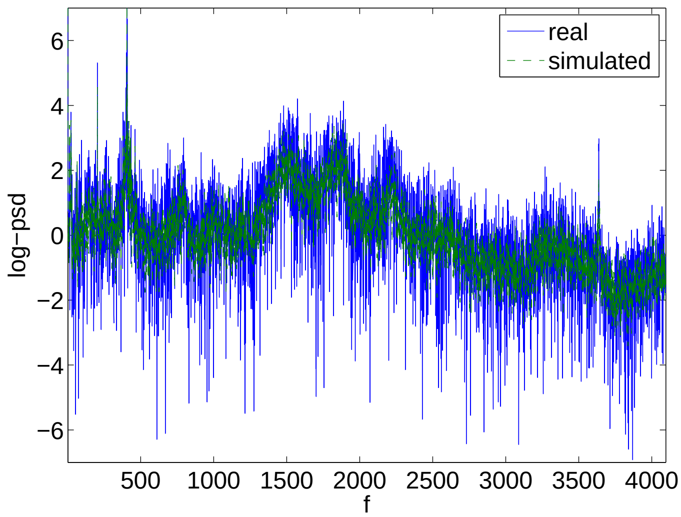


Figure 4: Randomly generated log-periodogram, conditionally on IMS bearing learning dataset.

following articles.

4.5. Results: Snecma turbofan engine

Results with the Snecma dataset are limited to random log-periodogram generation, since no failure data is available at the moment. Fig. 6 shows a good agreement between the learning set and random periodograms. This by-product of Bayesian analysis will be of great importance to test other algorithms by Monte-Carlo methods.

5. Conclusion and perspectives

The aim of this article was to show that in the domain of vibration health monitoring, when small datasets only are available because of industrial constraints, even with no mechanical knowledge, Bayesian modelling of healthy log-periodograms in the wavelet domain was a good strategy that ensures fault detection.

We have presented several algorithms inspired by statistical novelty detection, and by signal processing, in order to model healthy vibrations, to generate random spectra, then to detect failures.

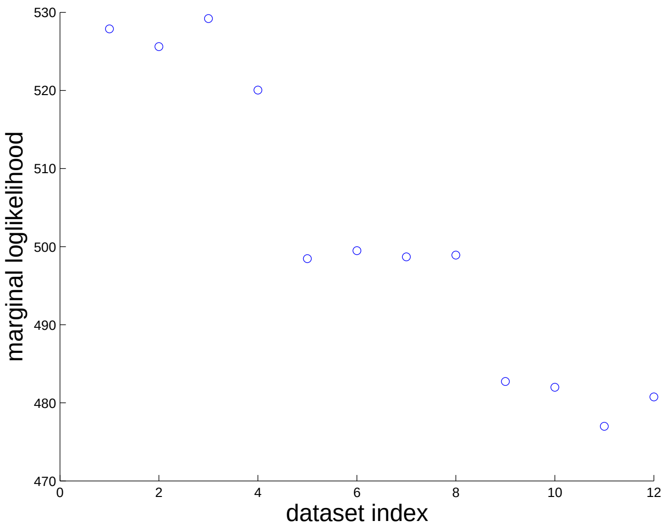


Figure 5: Marginal likelihood of log-periodograms (left) in healthy condition; (middle) partly deteriorated and (right) deteriorated.

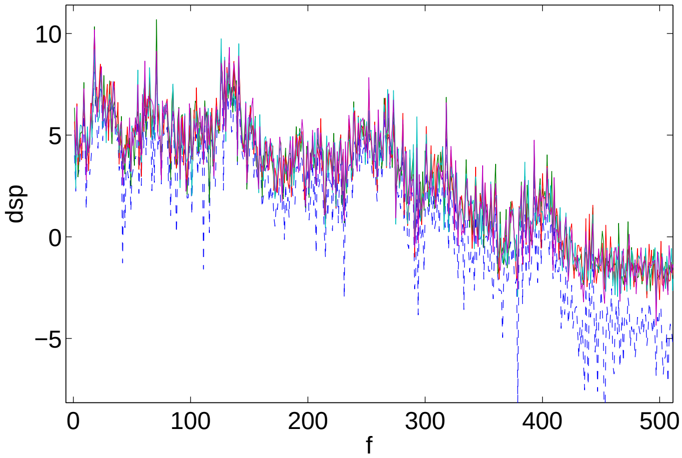
We have presented real illustrations of these algorithms, with deteriorated bearings recordings, and aircraft engines. This proves that Bayesian modelling of healthy periodograms in the wavelet domain ensures failure detection even with heavy constraints.

However, we faced several unexpected findings while running our experiments:

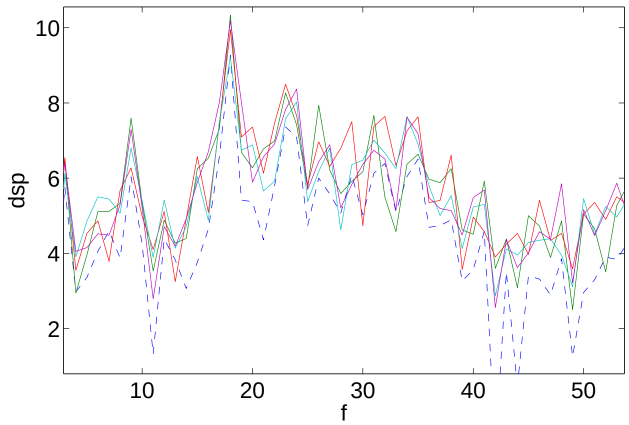
- the sparsity model widely used in the litterature for wavelet denoising may not be adapted when data are successive time series. If a given coefficient has a high value across several successive snapshots, it is very unlikely that the histogram of the measured coefficient is a mixture with a Dirac mass centered in zero.
- the updating scheme for the variance of posterior gives poor results in high scales.
- the use of marginal likelihood is questionnable if the model we obtain is not very accurate.

The following perspectives will be dealt with in future articles:

- a comparaison with state-of-the art methods will be established.



(a)



(b)

Figure 6: Randomly generated log-periodograms, conditionally on SNECMA aircraft engine learning dataset, at various zoom levels.

- in the spirit of Clifton et al. (2008) that apply Bayesian novelty detection to low-rank multiples of shaft speed, we plan to model the max of wavelet coefficients for the whole log-periodogram.
- hyperparameters may be learnt in a full bayesian way, i.e. with priors set on the fitted model.
- thresholding theory may be used to simplify the wavelet coefficients models.

Appendix A. Posterior

Here we give the expressions of the standard deviations that appear in the posterior eq.(14). σ_2 is the variance of the prior of the wavelet coefficient, which value is set according to eq. (12), which depends on j. We omit j subscripts for clarity:

$$\sigma_1 = \frac{1}{\sqrt{n}}\sqrt{\frac{\pi^2}{6}} \tag{A.1}$$

$$\frac{1}{\sigma_0} = \frac{1}{\sigma_1} + \frac{1}{\sigma_2} \tag{A.2}$$

where n is the number of samples.

References

Abramovich, F., Sapatinas, T., Silverman, B., 1998. Wavelet thresholding via a bayesian approach. J. R. Statisc. Soc.B 60 (4), 725–749.

Antoni, J., Bonnardot, F., Raad, A., Badaoui, M. E., 2004. Cyclostationary modelling of rotating machine vibration signals. Mechanical Systems and Signal Processing 18 (6), 1285 – 1314.
URL `http://www.sciencedirect.com/science/article/B6WN1-49S7XV6-1/2/d3fe040262d7df62b61303b7695bbfb5`

Bishop, C., 2006. Pattern Recognition and Machine Learning. Springer.

Brillinger, D., 1981. Time Series Data Analysis and Theory. Holden Day, San Francisco.

Clifton, D., 2006. Condition monitoring of gas-turbine engines. Tech. rep., Department of Engineering Science, University of Oxford.

Clifton, D., McGrogan, N., Tarassenko, L., S., K., Anuzis, P., D., K., 2008. Bayesian extreme value statistics for novelty detection in gas-turbine engines. In: Proc. of IEEE Aerospace. Montana, USA, pp. 1–11.

Clifton, D., Tarassenko, L., 2009. Novelty detection in jet engine vibration spectra. IET Condition Monitoring, –.

Kharyton, V., 2009. Fault detection of blades in blades of an aviation engines in operation. Ph.D. thesis, Ecole Centrale de Lyon.

Mallat, S., 2009. A wavelet tour of signal processing, 3rd ed. Elsevier.

Markou, M., Singh, S., December 2003a. Novelty detection: a review. part 1: statistical approaches. Signal Processing 83 (12), 2481–2497.

Markou, M., Singh, S., December 2003b. Novelty detection: a review. part 2: neural network based approaches. Signal Processing 83 (12), 2499–2521.

Moulin, P., Nov. 1994. Wavelet thresholding techniques for power spectrum estimation. IEEE Transactions on Signal Processing 42, 3126–3136.

Orsagh, R., Sheldon, J., Klenke, C., 2003. Prognostics/diagnostics for gas turbine engine bearings. In: Proceedings of IEEE Aerospace Conference. pp. –.

Peng, Z. K., Chu, F. L., 2004. Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography. Mechanical Systems and Signal Processing 18 (2), 199 – 221. URL `http://www.sciencedirect.com/science/article/B6WN1-491HXGB-2/2/0268975d2594e5b67e35f8b9e93cc41a`

Peng, Z. K., Chu, F. L., Tse, P. W., 2005. Detection of the rubbing-caused impacts for rotor-stator fault diagnosis using reassigned scalogram. Mechanical Systems and Signal Processing 19 (2), 391 – 409. URL `http://www.sciencedirect.com/science/article/B6WN1-4B4288R-1/2/a45cae5986f6e70c6fad364f442a8011`

Pensky, M., Vidakovic, B., De Canditiis, D., 2007. Bayesian decision theoretic scale-adaptive estimation of a log-spectral density. Statistica Sinica 17, 635–666.

Percival, D., Walden, A., 2000. Wavelets methods for Time Series analysis. CUP.

Qiu, H., Lee, J., Lin, J., 2006. Wavelet filter-based weak signature detection method and its application on roller bearing prognostics. Journal of Sound and Vibration 289, 1066–1090.

Robert, C., Casella, G., 2010. Introducing Monte Carlo methods with R. Springer.

Tarassenko, L., Nairac, A., Townsend, N., Cowley, P., 1999. Novelty detection in jet engines. In: Colloqium on Condition Monitoring, Imagery, External Structures and Health. IEEE, pp. 41–45.

Vidakovic, B., 1999. Statistical Modeling by Wavelets. Wiley.

Wang, W., Ismail, F., Golnaraghi, M., 2001. Assessment of gear damage monitoring techniques using vibration measurements. Mechanical Systems and Signal Processing 15 (5), 905 – 922.
URL http://www.sciencedirect.com/science/article/B6WN1-456JSBF-H/2/d49603d4d46addef88d9e254d53421ee

Ypma, A., Duin, R., 1997. Novelty detection using self-organizing maps. In: Kasabov, N., Kozma, R., Ko, K., O'Shea, R., Goghill, G., Gedeon, T. (Eds.), Progress in Connectionist-Based Information Systems, Vol.II. Springer Verlag, pp. 1322–1325.