Blind vibration filtering using envelope spectrum indicators for bearing and gear fault detection without knowledge of machine kinematics

Cédric Peeters, Jérôme Antoni, Jan Helsen

Blind vibration filtering using envelope linear prediction for fault detection without knowledge of machine kinematics

Cédric Peeters[1], Jérôme Antoni[2], Jan Helsen[1]

[1]Vrije Universiteit Brussel
Pleinlaan 2, 1050 Brussel, Belgium
cedric.peeters@vub.be

[2]LVA, INSA de Lyon,
F-69621, Villeurbanne cedex, France

Abstract

The central idea behind this paper is to propose a means to filter out vibration signals of interest from a fault detection perspective without actually having knowledge about the kinematics of the machine. In other words, this paper investigates blind filters that do not require a-priori knowledge about the fault frequencies, e.g. of a bearing or gear. This kind of approach opens the door for the condition monitoring of complex machines where insufficient information is available about the inner components or where replacements have been carried out that changed characteristic frequencies and that were not logged. This feat is achieved by employing the squared envelope as a metric for the blind filter. The main assumption of the proposed method is that when a fault occurs, it introduces a second-order cyclostationary (CS2) component in the vibration signal which manifests itself in the squared envelope (SE) as a harmonic sine modulation at its corresponding fault frequency. This modulation correspondingly also increases the sparsity of the envelope spectrum. To avoid interfering influences of CS1 components, the signal is typically pre-whitened, e.g. through linear prediction filtering, cepstrum editing, etc. The paper investigates the minimization of the relative prediction error of the linear prediction of the squared envelope for use in the iterative updating procedure of the blind filter.

Keywords

Blind filtering, vibrations, linear prediction, sparsity, envelope, fault detection

1 Introduction

Complex machines nowadays can consist of dozens of bearings and gears, with modern examples being the gearboxes of wind turbines and helicopters. These gearboxes typically have one or multiple planetary gear stages in combination with parallel gear stages. Not all kinematic information about the system might be available to the machine operator, or the information might be inaccurate due to reparations with new components. This issue constitutes the need for a method capable of tracking the condition of these components without the need for a-priori knowledge about the kinematics.

One of the most popular approaches for fault detection is to look at the cyclostationary behavior of the vibration signal [1, 2, 3, 4, 5, 6, 7, 8, 9]. Inspired by this fact, this paper investigates the possibility to utilize the cyclostationary content of a signal in a blind manner. Therefore, instead of just looking at the statistics of the time waveform, the squared envelope of the signal is employed as a means to gain more information about potential defects. From experience it is known that most mechanical faults of bearings or gears induce some form of cyclostationary behavior in the observed vibration signals [4, 10, 11]. This cyclostationary behavior alters the modulation signature of the signal. For example a repetitive impulse train (similar to a bearing fault) introduces harmonics at the repetition frequency of the impulses into the envelope. This means that the envelope signal becomes more predictable and can thus be fitted with an autoregressive model. This property is thus exploited to find a filter that minimizes the relative prediction error of the squared envelope since it is assumed that a good fit corresponds to a mechanical fault and not to normal behavior.

An important remark about the proposed approach is that the blind filtering methodology described in this paper cannot be categorized as blind deconvolution, blind signal separation, or denoising. The proposed approach namely does not attempt to deconvolve the signal in order to recover the source signal (e.g. impulses), nor does it attempt to separate signals from a mixture or remove noise from the signal without distorting it. In fact, it actually does distort the signal such that the squared envelope is as predictable as possible. This is a fairly new concept since from this perspective the algorithm does not care about restoring the signal or recovering the source signals. Instead the algorithm just tries to maximize the figure of merit and thus enhance the envelope spectrum. It is important to take into account this distinction when inspecting filtering results since the results might not correspond to what is expected.

This paper attempts to highlight the utility of blind filtering based on the signal envelope and the versatility of the Rayleigh quotient regarding the indicator choice for the blind filtering step. First, the theoretical background is explained in Section 2. The indicator choice and the derivation of the Rayleigh quotients are described. Next, the method is validated on simulated signals in Section 3 and experimental data of a gearbox data set in Section 4. The results show that the proposed approach is capable of extracting a cyclostationary fault signature and that the prediction error measure of the envelope in itself can be used as a tracking parameter.

2 Methodology

The idea of the proposed methodology is to exploit the predictability of a fault modulation signature by trying to fit it with a linear prediction (LP) filter or auto-regressive (AR) filter. The prediction error of an auto-regressive all-poles model of the squared envelope serves then as the metric of interest. The better the AR model can fit the actual envelope, the more predictable and thus the less noisy it is. This means that if there is a signal component present with e.g. a clean sinusoidal amplitude modulation, the AR model is then capable of predicting future samples accurately which in turn corresponds to a low prediction error. This does indicate again the need for prewhitening the signal to make sure the AR model does not try to fit the envelope of deterministic components in the signal.

2.1 Blind filtering

The concept of blind filtering is to find a filter that maximizes a certain criterion of the signal starting from a noisy measured signal $\mathbf{x}$:

$$\mathbf{s} = \mathbf{x} * \mathbf{h} \tag{1}$$

where $\mathbf{s}$ is the estimated input, $\mathbf{h}$ is the inverse filter, and $*$ refers to the convolution operation. It should be noted that vectors and matrices are set in bold font to illustrate the difference with scalars. The convolution is expressed as:

$$\mathbf{s} = \mathbf{X}\mathbf{h} \tag{2}$$

$$\begin{bmatrix} s_{N-1} \\ \vdots \\ x_{L-1} \end{bmatrix} = \begin{bmatrix} x_{N-1} & \cdots & x_0 \\ \vdots & \ddots & \vdots \\ x_{L-1} & \cdots & x_{L-N-2} \end{bmatrix} \begin{bmatrix} h_0 \\ \vdots \\ h_{N-1} \end{bmatrix}$$

with L and N the number of samples of $\mathbf{s}$ and $\mathbf{h}$ respectively.

Now the squared envelope $\varepsilon_{\mathbf{x}}$ can be defined as follows:

$$\varepsilon_{\mathbf{x}} = |\mathbf{s}|^2 = |\mathbf{X}\mathbf{h}|^2 \tag{3}$$

It can also be written as:

$$\varepsilon_{\mathbf{x}} = \begin{bmatrix} s_0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & s_{L-N+1} \end{bmatrix}^H \mathbf{X}\mathbf{h} = diag(\mathbf{s^H})\mathbf{X}\mathbf{h} \tag{4}$$

with $\mathbf{s^H}$ being the Hermitian transpose of $\mathbf{s}$, and $diag(\mathbf{s^H})$ being a diagonal matrix with the values of the vector $\mathbf{s^H}$ on its diagonal.

Figure 1: Linear prediction coding.

2.2 Derivation of LP-envelope filter

The relative prediction error of the AR model is closely related to the spectral flatness as the AR model also maximizes the spectral flatness of the squared envelope prediction error [12]. The relative prediction error (RPE) of the AR model of the squared envelope is given by:

$$RPE = \frac{\sigma_e^2}{\sigma_{SE}^2} \tag{5}$$

with σ_e being the prediction error of the squared envelope, and σ_{SE} being the standard deviation of the squared envelope. The autoregressive coefficients can be obtained by fitting a linear prediction model on the squared envelope. The standard LPC representation of a signal $x(n)$ for a model of order N is:

$$x(n) = \sum_{i=1}^{N} a_i x(n-i) + e(n) \tag{6}$$

with a_i the autoregressive coefficients, and where the error $e(n)$ can be obtained by:

$$e(n) = \sum_{i=0}^{N} a_i x(n-i) \ with \ a_0 = 1. \tag{7}$$

Figure 1 illustrates the straightforward process. The autoregressive filtering necessary to obtain the prediction error $\mathbf{e}$ of the squared envelope ε_x can be written in matrix notation as follows:

$$\mathbf{e} = \mathbf{A} diag(\mathbf{s^H})\mathbf{s} \tag{8}$$

with $\mathbf{A}$ a band matrix containing the autoregressive coefficients as follows:

$$\mathbf{A} = \begin{bmatrix} a_0 & a_1 & \dots & a_N & 0 & \dots & 0 \\ 0 & a_0 & a_1 & \dots & a_N & \ddots & \vdots \\ & & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & a_0 & a_1 & \dots & a_N \\ & & & & \ddots & \ddots & \vdots \\ & & & & & a_0 & a_1 \\ 0 & & & \dots & & 0 & a_0 \end{bmatrix} \tag{9}$$

Since the intent is actually to minimize the relative prediction error in Eq. 5 and not maximize it, the ratio to be maximized is actually the inverted RPE. Equation 5 can thus be written in the following manner:

$$\boxed{\frac{\sigma_{SE}^2}{\sigma_e^2} = \frac{{\varepsilon_\mathbf{x}}^H \varepsilon_\mathbf{x}}{\mathbf{e}^H \mathbf{e}} = \frac{\mathbf{h^H X^H} diag(\mathbf{s}) diag(\mathbf{s^H})\mathbf{Xh}}{\mathbf{h^H X^H} diag(\mathbf{s})\mathbf{A}^H \mathbf{A} diag(\mathbf{s^H})\mathbf{Xh}} = \frac{\mathbf{h^H R}_{XW_1X}\mathbf{h}}{\mathbf{h^H R}_{XW_2X}\mathbf{h}}} \tag{10}$$

The generalized Rayleigh quotient [13] can be recognized in Eq. 10 and can be maximized using an iterative maximization of the eigenvalues:

$$\lambda = \frac{\mathbf{h^H R}_{XW_1X}\mathbf{h}}{\mathbf{h^H R}_{XW_2X}\mathbf{h}} \tag{11}$$

The Rayleigh quotient has the interesting property that its maximal value with respect to $\mathbf{h}$ is equivalent to its largest eigenvalue λ and corresponding eigenvector. Thus, maximizing the Rayleigh quotient allows finding the maximal values of the corresponding indicator and filter. In order to obtain real eigenvalues however, the correlation matrices $\mathbf{R}_{XW_1X}$ and $\mathbf{R}_{XW_2X}$ need to be Hermitian, and $\mathbf{R}_{XW_2X}$ needs to be positive semidefinite. If these conditions are met, the Rayleigh quotient offers an efficient means to calculate iteratively the filter coefficients. Only the largest eigenvalue and corresponding eigenvector need to be computed in each iteration, which can be achieved efficiently by using algorithms such as the power method [14].

The generalized eigenvalue problem to be solved can be formulated as such:

$$\mathbf{R}_{XW_1X}\mathbf{h} = \mathbf{R}_{XW_2X}\mathbf{h}\lambda \tag{12}$$

The iterative algorithm used to minimize the prediction error consists out of four basic steps:

1. Assume an initial guess for **h**
2. Estimate $\mathbf{R}_{XW_1X}$ and $\mathbf{R}_{XW_2X}$ based on **h** and **X** using Eq. 10
3. Solve Eq. 12 to find λ_{max} and a new filter **h** that corresponds to a higher value of the used criterion
4. Return to step 2 using the new **h** until convergence is reached or the maximum number of iterations

The name of the proposed method is abbreviated to LPE (Linear Prediction of Envelope) to keep the text concise.

3 Simulation

To validate the proposed approach, a straightforward simulated case is first considered. To add some point of reference, the performance of the proposed LPE method is compared to Minimum Entropy Deconvolution (MED), which is a commonly used blind deconvolution filtering technique to maximize the kurtosis. An outer race bearing fault signal is simulated with a normalized sample rate of 1 Hz and duration of 20000 samples. White Gaussian noise is added at varying degrees of signal-to-noise ratio. The parameters used for the simulation of the bearing fault signal:

- the outer race bearing fault frequency, $f_{BPFO} = 0.31Hz$
- the resonance frequency excited by the fault, $f_{IRF} = 0.25Hz$
- the damping ratio, $\zeta = 0.1$
- the jitter, $J = 1\%$

The signal-to-noise ratio of the fault signal is varied linearly from -30 dB to + 15 dB. The overall variance of the signal is kept constant however. The filter length is chosen to be 15 and 150 samples respectively for the proposed LPE method and MED filtering. The max number of iterations is set at 100.

Figure 2 shows the evolution of the indicators after blind filtering using the proposed method and MED. It can be observed that the relative prediction error starts to decrease gradually around -17 dB SNR. In this case the proposed method outperforms MED mainly because the generated bearing fault signal does not immediately give rise to a large increase in the kurtosis of the signal. This is due to the fault frequency being relatively high and therefore causing the exponentially decaying impulse responses to smear over. This is not an ideal scenario for kurtosis maximization. However, when inspecting the envelope, the fault frequency modulation shows up faster because it does not suffer as much from this smearing. Instead it detects still the envelope fluctuation at exactly the fault frequency, albeit less pronounced than it would be without smearing.

Figure 3 shows a color map in grayscale of the evolution of the normalized squared envelope spectra. The fault frequency of 0.31 Hz can be clearly distinguished after approximately -17 dB and 2 dB SNR for respectively the proposed and MED method. An alternative 3D view of this color map is provided in Figs. 4 and 5.

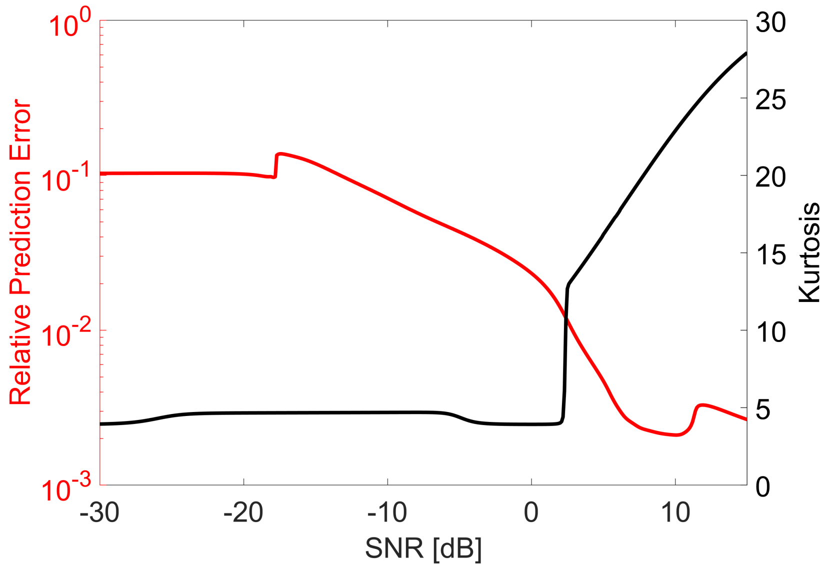


Figure 2: Trending of Relative Prediction Error (RPE) versus kurtosis after blind filtering using the proposed LPE approach and MED respectively.

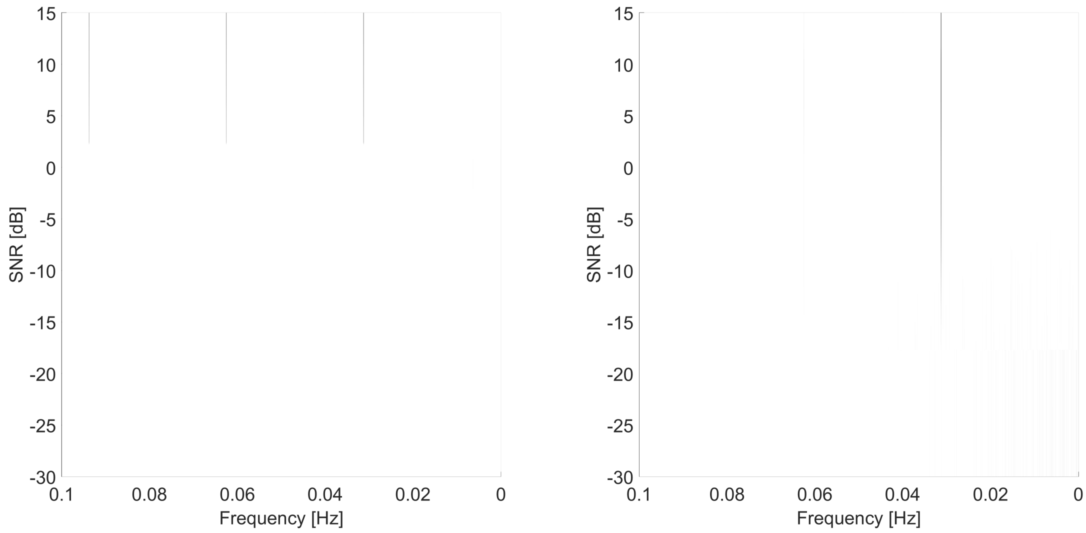


Figure 3: Color maps (in grayscale) of the squared envelope spectra after blind filtering using the MED (left) and proposed approach (right) respectively. The black lines appearing at approximately -17 dB and 2 dB SNR for respectively the proposed and MED method correspond to the 0.31 Hz fault frequency.

4 Experimental application

To verify whether the proposed method also works for more complex signals, the approach is tested on an experimental vibration signal. The real-world case chosen for this purpose is the IMS bearing prognostic dataset [15]. This dataset contains an outer race bearing fault in measurement campaign 2. The BPFO is approximately 236 Hz. The record chosen for filtering is n 690 since it is known that there is already damage distinguishable using other signal processing methods but the damage is not distinguishable directly from the raw vibration signal. The result after filtering using the LPE method is displayed in Fig. 6. While the original input signal does not show any clear signs of damage, the filtered signal evidently exhibits a strong modulation at exactly the BPFO of 236.1 Hz and its second harmonic at 472 Hz. This is a favorable result since no a-priori knowledge of the fault frequency was provided to method. It should also be noted that in testing the MED filtering method failed to extract any clear indication of the fault for the same signal.

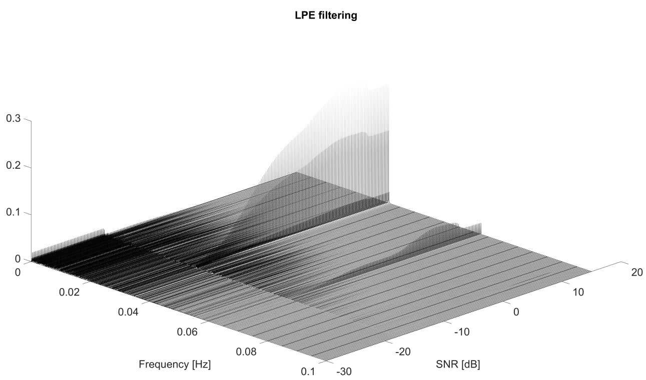


Figure 4: Squared envelope spectrum waterfall plot after filtering with the proposed method based on the relative prediction error.

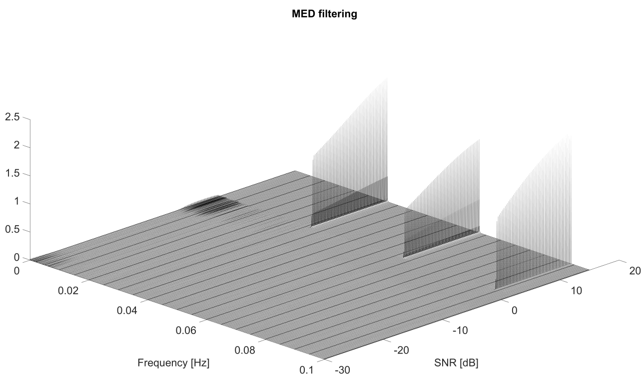
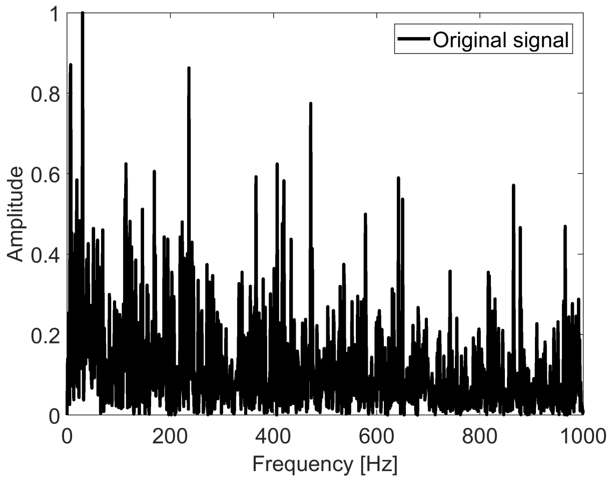


Figure 5: Squared envelope spectrum waterfall plot after filtering with the MED method based on kurtosis.



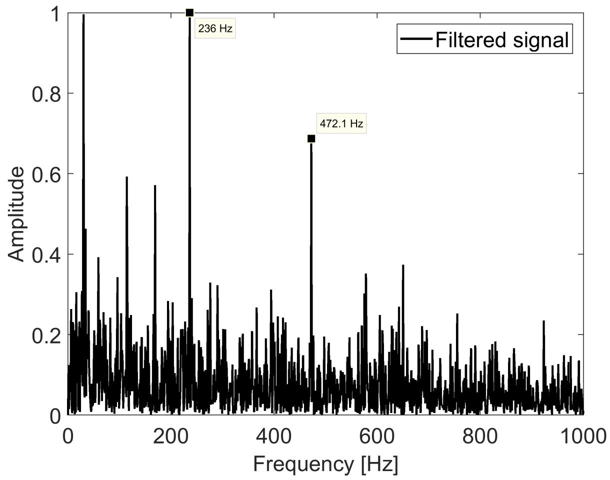


Figure 6: Experimental case: (Left) The envelope spectrum of the raw input signal. (Right) The envelope spectrum after filtering with the proposed LPE method.

5 Conclusions

This paper describes a novel fault detection approach to blind filtering of vibration signals. Instead of looking to maximize a statistic of the time waveform, it proposes to utilize the squared envelope and more in particular the predictability of it. Most fault signatures of bearing or gear faults induce structured and thus predictable second-order cyclostationary behavior. The main assumption is thus that an increase in the envelope predictability is linked to the emergence of a fault. This knowledge is then used to derive a blind filtering approach employing a generalized Rayleigh quotient iteratively to optimize the filter coefficients. The proposed method is validated on both simulation data as experimental data. Both scenarios prove that the proposed approach can work in an efficient and accurate manner. Employing the envelope instead of the time waveform, as done by MED filtering, also alleviates some of the limitations the latter can have. An example of such a limitation is the fact that MED filtering assumes that the fault signal has a high kurtosis value due to its impulsiveness. In reality this is not always the case. Additionally, MED filtering has the tendency to deconvolve a single high amplitude peak due to the fact that kurtosis is maximized in such a scenario. This is unlikely for a rotating component where impulses are generated at every rotation. Therefore the use of the envelope signal is a logical way to assess the presence of faults in rotating machinery.

While there is no need to input the fault frequency in the method, there is still one parameter that can influence the outcome, namely the filter length. A general recommendation for the filter length is to try to keep it as short as possible. Not only does this reduce the computation time, but it also means that there are less filter coefficients that need to be optimized and thus less coefficients that vary. The updating procedure also becomes more stable and fluctuates less due to numerical inaccuracies (e.g. high condition numbers of the matrices).

The results presented in this paper are promising as they prove that detailed information about the system of interest is not always necessary to still perform health tracking. It also opens the door for more and different approaches using blind filtering based on the envelope and its properties instead of using the raw time waveform.

References

[1] Zimroz R and Bartelmus W 2009 Gearbox condition estimation using cyclo-stationary properties of vibration signal *Key Engineering Materials* vol 413 (Trans Tech Publ) pp 471–478

[2] Antoni J, Daniere J and Guillet F 2002 *Journal of sound and vibration* **257** 815–837

[3] Randall R B 2011 *Vibration-based condition monitoring: industrial, aerospace and automotive applications* (John Wiley & Sons)

[4] Antoni J, Bonnardot F, Raad A and El Badaoui M 2004 *Mechanical systems and signal processing* **18** 1285–1314

[5] McCormick A and Nandi A 1998 *Mechanical systems and signal processing* **12** 225–242

[6] Antoniadis I and Glossiotis G 2001 *Journal of sound and vibration* **248** 829–845

[7] Capdessus C, Sidahmed M and Lacoume J 2000 *Mechanical systems and signal processing* **14** 371–385

[8] Borghesani P, Pennacchi P, Ricci R and Chatterton S 2013 *Mechanical Systems and Signal Processing* **40** 38–55

[9] Antoni J, Xin G and Hamzaoui N 2017 *Mechanical Systems and Signal Processing* **92** 248–277

[10] Randall R B, Antoni J and Chobsaard S 2001 *Mechanical systems and signal processing* **15** 945–962

[11] Randall R B and Antoni J 2011 *Mechanical Systems and Signal Processing* **25** 485–520

[12] Stoica P, Moses R L *et al.* 2005

[13] Hom R A and Johnson C R 1985 *Cambridge University Express*

[14] Mises R and Pollaczek-Geiringer H 1929 *ZAMM-Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik* **9** 152–164

[15] Lee J, Qiu H, Yu G and Lin J 2007 *Moffett Field, CA: IMS, Univ. Cincinnati. NASA Ames Prognostics Data Repository, NASA Ames*