Refining Envelope Analysis Methods using Wavelet De-Noising to Identify Bearing Faults

Edward Max Bertot[1], Pierre-Philippe Beaujean[2], and David Vendittis[3]

[1,2,3] *Florida Atlantic University, Boca Raton, FL, United States*
ebertot@fau.edu
pbeaujea@fau.edu
dvendittis@aol.com

ABSTRACT

In the field of machine health monitoring, vibration analysis is a proven method for detecting and diagnosing bearing faults in rotating machines. One popular method for interpreting vibration signals is envelope-demodulation, which allows the maintainer to clearly identify an impulsive fault source and its severity. In some cases, in-band noise can make impulses associated with incipient faults difficult to detect and interpret. In this paper, we use Wavelet De-Noising (WDN) after envelope-demodulation to improve the accuracy of bearing fault diagnostics. This contrasts the typical approach of de-noising raw vibration signals prior to demodulation. We find that WDN removes low amplitude harmonics and spurious reflections which may interfere with FFT techniques to identify low-frequency peaks in the signal spectrum. When measuring impact frequencies in the time-domain using a peak-thresholding method, the proposed algorithm exhibits increasingly confident periodicity at bearing fault frequencies.

1. INTRODUCTION

1.1. Bearing Fault Diagnosis

A faulty bearing will typically create periodic, impulsive vibrations, which are proportional to rotational speed. These vibrations may be recorded and analyzed to reveal the nature of a given fault. Systems with multiple bearings and gear reduction systems will exhibit unique fault frequencies due to varying component dimensions and operating speeds. This simple observation may be exploited to determine exactly which component is failing (Qui, Lee, Lin, & Yu, 2006). In more sophisticated systems, multiple sensors are often used to indicate fault locations based on local vibration power levels (Waters & Beaujean, 2013).

1.2. Envelope Analysis

Within a given structure, fault-induced impulses will amplitude-modulate mechanical resonances (McFadden & Smith, 1984). Research on which this paper is based (Waters & Beaujean, 2013) utilizes envelope analysis to extract impulses from the modulated signal, which allows for quick diagnosis of apparent mechanical problems.

However, incipient faults are rather difficult to detect using this method, due to lower signal-to-noise ratio (SNR). Extraneous noise sources such as nearby modal resonances, vibrational reflections, and vibrational harmonics corrupt the envelope signal. We find that low SNR degrades early-detection abilities and in turn deteriorates estimates of Remaining Useful Life (RUL). These noise sources are in-band and non-white, so their removal is less than trivial.

To combat a low SNR in the demodulated signal, we require a "de-noising" technique. This research focuses on wavelet de-noising and its use in vibration analysis, particularly as a post-processing scheme for envelope analysis. A secondary objective is to reduce user-interaction with the algorithm's parameters to obtain beneficial results.

1.3. Wavelet De-Noising

Many techniques have been devised for noise removal via signal processing. For our purposes, the algorithm must process non-stationary signals with good time-resolution. Vibration statistics will be in constant flux, given changes in bearing wear, speed, and operating environment. More importantly, it must perform without *a priori* knowledge of the noise.

As far as the aforementioned requirements specify, Wavelet De-Noising (WDN) is a proven candidate. The wavelet transform outperforms the Short-Time Fourier Transform (STFT) in terms of temporal resolution, allowing it greater flexibility in analyzing non-stationary signals (Rioul & Vetterli, 1991). It has also been demonstrated that WDN requires no knowledge of the noise level in order to optimally remove it (Donoho, 1995).

1.4. Paper Structure

Section 2 provides a brief overview of wavelet de-noising and its function, and reviews previous literature pertaining to PHM applications. Section 3 explains the proposed methodology, then sections 4 and 5 contain results supporting the use of WDN to help interpret demodulated vibration signals, and Section 6 contains a few concluding remarks.

2. BACKGROUND

2.1. Discrete Wavelet Transform

A more in-depth discussion of wavelet techniques can be found in (Daubechies, 1992). The wavelet transform, given as the operator W, is easily visualized in the continuous domain:

$$W_f(a,b) = \int_{-\infty}^{\infty} f(t)\psi_{a,b}(t)dt \tag{1}$$

where f is an arbitrary function of the independent variable t, and $\psi_{a,b}$ is a family of wavelet functions defined by scaling and shifting – respectively a and b,

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right) \tag{2}$$

where ψ is a prototype function, or wavelet kernel.

In order to utilize this transform for sampled data, we discretize the scaling and shifting parameters in the following manner:

$$a_m = a_0^m \tag{3}$$

$$b_{m,n} = nb_0 a_0^m \tag{4}$$

where m and n are the discrete analogues of frequency and time, respectively.

Notably, the shifting parameter b is a function of scale a. This illustrates a crucial advantage of the Discrete Wavelet Transform (DWT); the distribution of information in frequency is dyadic, or octave-band. For analyzing natural signals, this is highly useful (Rioul & Vetterli, 1991).

The DWT results in a set of wavelet coefficients d, which are given by the inner product

$$d_{m,n} = \langle f(t), \psi_{m,n}(t) \rangle \tag{5}$$

which, when the proper wavelet family is chosen, represents

a frequency-orthogonal decomposition of the original signal into subbands which are logarithmically spaced in frequency, as shown in Figures 1 and 2. In Figure 1, the wavelet inner product is functionally equivalent to BPF and LPF, or bandpass and lowpass filtering.
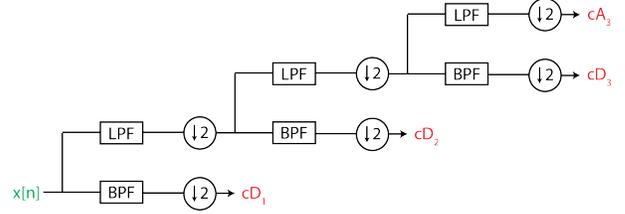
Figure 1. Octave subband tree structure with three levels of decomposition. Each filtering results in a set of coefficients, typically referred to as detail (high frequency, cD) and approximation (low frequency, cA) coefficients. If this pattern is repeated until cD_6 and cA_6, the 6-level decomposition shown in Figure 2 will result.
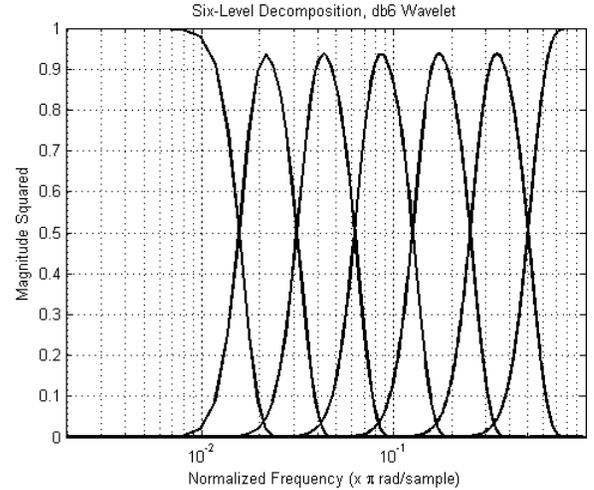
Figure 2. Filter magnitude responses of a six level wavelet decomposition, using the db6 wavelet. Note the logarithmic frequency scale.

2.2. Coefficient Thresholding

As originally proposed in (Donoho & Johnstone, 1994), the linear soft thresholding function is given by

$$\tau(x) = \begin{cases} x - \lambda \mathrm{sgn}(x), & |x| \geq \lambda \\ 0, & |x| < \lambda \end{cases} \tag{6}$$

where x are the values being thresholded, $\mathrm{sgn}(x)$ is the sign of x, and λ is the threshold below which values are set to zero. Donoho and Johnstone (1994) prove that the threshold λ for near-optimality (in the minimax sense) is calculated as

$$\lambda = \sigma_x \sqrt{2\log(N)} \tag{7}$$

where N is the number of samples in the time series and σ_x is the noise deviation. Exact noise statistics are difficult to estimate without *a priori* characteristics or reference measurements. A simplifying assumption is to consider Gaussian noise as the dominant source in an incipient fault situation, as suggested in Bozchalooi and Liang (2007). Therefore, the noise deviation $sigma_x$ is just the unbiased estimate of the standard deviation of the input signal.

$$\sigma_x \approx \sqrt{\frac{1}{N-1}\sum_{k=1}^{N}(x_k - \mu_x)^2} \qquad (8)$$

where x_k are sample values and μ_x is the arithmetic mean of the time series.

When the thresholding function is applied to orthogonally derived wavelet coefficients, the result is a de-noised version of the original signal.

2.3. Existing Literature

Qui et al. (2006) discussed wavelet domain techniques for vibration analysis applications. The authors use the same method described in Section 2, but they criticize the use of WDN for vibration signals due to tuning difficulties:

> [...] there are other factors influencing the effectiveness of [wavelet] de-noising, such as the wavelet decomposition level and threshold rescaling method selection, which make the de-noising problem even more intricate. Since there are no explicit guidelines for how to tune the existing parameters, most of the time de-noising becomes a trial-and-error process. (Qui et al., 2006, pg. 1080)

There is much truth to these statements, and using WDN on raw vibrational signals generally gives unpredictable results. However, this paper concludes that WDN is quite functional in the context of envelope-demodulated vibration signals.

3. SIGNAL FLOW & METHODOLOGY

Typically, de-noising algorithms are used as a pre-processing step to improve the effectiveness of subsequent signal processing. However we find that when used prior to envelope demodulation, WDN removes low-amplitude modal resonances that allow the Hilbert Transform to work well. If the de-noising is performed after demodulation, the impulse signal is more effectively de-noised.

The full signal processing procedure is as follows:

1. The raw vibration signal is Hilbert filtered at a chosen modal vibration frequency, resulting in a bandpass signal.
2. A Hilbert transform is performed, bringing the signal into the baseband.

3. WDN is used to attenuate lower amplitude harmonics and vibrational reflections.
4. The signal is searched for faults using peak detection in both time and frequency.

This report mainly focuses on the third step of this process.

3.1. Time-Domain Detection

For time-domain peak detection, the MATLAB ® function `findpeaks` is used to find local maxima. These peaks are thresholded at thr_e, which is a function of the average signal power,

$$thr_e = \alpha \frac{1}{N_e}\sum_{i=0}^{N_e-1} e_i^2. \qquad (9)$$

where N_e is the number of samples in the envelope signal e. The constant α allows for adjustment to this threshold. This function will remove smaller peaks that are not associated with larger impacts.

The times between all successive peaks in the envelope signal are measured, resulting in a vector of impulse periods. The inverse of this vector is a set of impulse frequencies. A histogram will reveal higher concentrations on fault frequencies.

3.2. Frequency-Domain Detection

We use a Welch PSD estimate to visualize the distribution of energy in the frequency domain. This allows for smaller time windows and reduces spurious peaks in the FFT via averaging.

4. SYNTHETIC SIGNAL TESTING

4.1. Setup

A synthetic vibration signal was constructed to test WDN effectiveness on a controlled envelope signal.

$$d[n] = e^{-\gamma\tau}sin(\omega\tau) \qquad (10)$$

$$\tau = \frac{n}{f_s} \qquad (11)$$

n is the sample number, f_s is the sampling frequency, τ is time relative to $n = 0$, and ω is the simulated modal resonance frequency (rad/sec). γ is the exponential decay constant. This damped sine function is windowed and repeated in time to simulate a periodic impact, much like a bearing fault may produce in rotating machinery.

Modeled after real fault signals from the Case Western Reserve University bearing data set (*Bearing Data Center*, 2013), these values are inferred by observing real signals:

$$\gamma \approx 1000 \quad \omega \approx 5000\pi\frac{rad}{sec} \quad f_s = 48\text{kHz}$$

White noise is added to the signal at $\text{SNR}_T \approx 0\text{dB}$, where SNR_T is the time-domain signal-to-noise ratio. This is calculated within a window of one time-constant of the exponentially decaying signal. This prevents inclusion of zeros between pulses, which would artificially reduce the SNR measure.

4.2. Discussion of Parameter Selection

Sensible parameter choices are derived from this experiment, which help to effectively de-noise the envelope signal.

The first parameter is n_d, or the number of decomposition levels. Selection of this value essentially determines the bandwidth of the lowest two subbands. If the desired signal is placed between subbands, then undesired attenuation may occur during thresholding.

For baseband envelopes, the number of decompositions depends on the highest possible fault frequency. In the case of a rolling element bearing, this is usually BPFI (Ball Pass Frequency of Inner raceway) (McFadden & Smith, 1984). Therefore, to determine the maximum number of decompositions allowable, we find n_d such that

$$f_s \frac{\pi/2^{n_d}}{2\pi} > \text{BPFI}. \qquad (12)$$

This will ensure that the frequencies of interest are not lost between subbands.

The other important parameter is the wavelet, ϕ, which will determine the amount of energy leakage between subbands. Higher-order wavelets decrease subband leakage, but require more computational power. In the time-domain, baseband envelopes simply correspond to a lowpass-filtered impulse train. In the frequency domain, this corresponds to high energy concentrations near DC. Higher order wavelets will more accurately de-noise and reconstruct the low frequency band, which contains frequencies of interest. Throughout these experiments, the Daubechies 20-tap wavelet (db20) is sufficient.

4.3. Results

The synthesized signal is Hilbert-filtered (bandpass) at ω, Hilbert transformed (demodulated), and WDN is applied. The waveforms in Figure 3 show all stages of the algorithm.

To de-noise the envelope signal, we choose to use 10 levels of decomposition. The reasoning, using Equation 12, is that 10 levels of decomposition will give a lowpass (scaling filter) cutoff at $\approx 24\text{Hz}$. This cutoff needs to be set above the synthesized fault frequency, which is 20Hz.

4.3.1. Frequency-Domain Detection

Figure 4 shows a low-frequency Welch PSD of the signal before and after WDN, with the expected fault frequencies
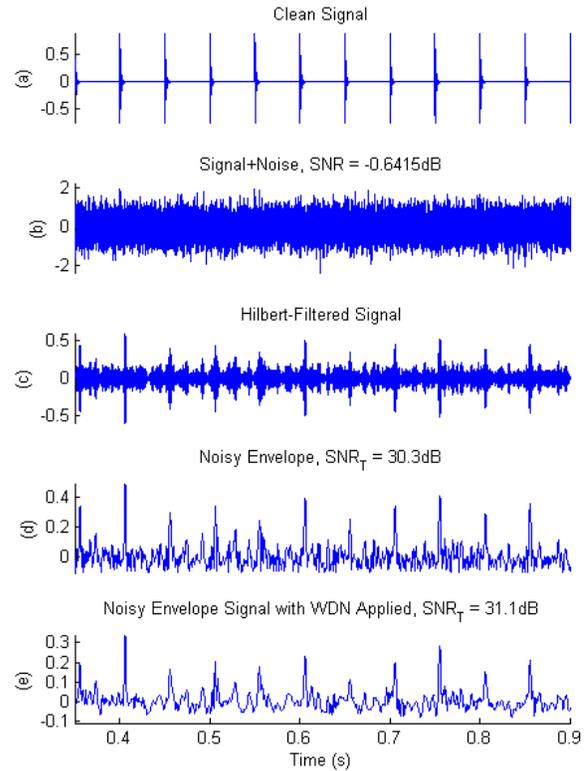
Figure 3. WGN added to damped sine pulses. The fourth and fifth plots show an increase in SNR_T by $\approx +1\text{dB}$ from applying WDN. The benefits of this procedure are not immediately obvious in the time-domain. Wavelet decomposition using db20 wavelet at a depth of 10 levels.

in grey. WDN removes higher harmonics that dominate the PSD, which increases the likelihood of proper fault identification using frequency-domain techniques.

4.3.2. Time-Domain Detection

To identify the fault in the time-domain, the signal is run through a peak detector and thresholded. The confidence interval plot in Figure 5 shows the improvement in detection ability for a wider range of α. The bands around the estimate denote 95% confidence intervals. The histograms in Figures 6 and 7 demonstrate what happens as α becomes too high, and the time-domain plot in Figure 8 shows the location of the threshold for $\alpha = 46$.

4.4. Remarks

The WDN algorithm successfully attenuates non-fault related envelopes in the signal, increasing the probability of proper fault identification using both frequency-domain (PSD) and time-domain (peak thresholding) methods. The confidence
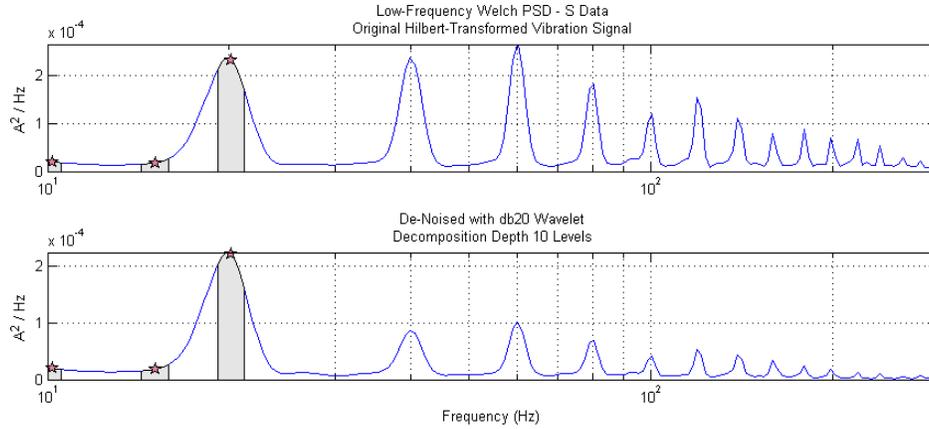
Figure 4. Welch PSD of signals shown in Figure 3 before (top) and after (bottom) WDN. Gray area shows $\pm 10\%$ of possible fault frequencies
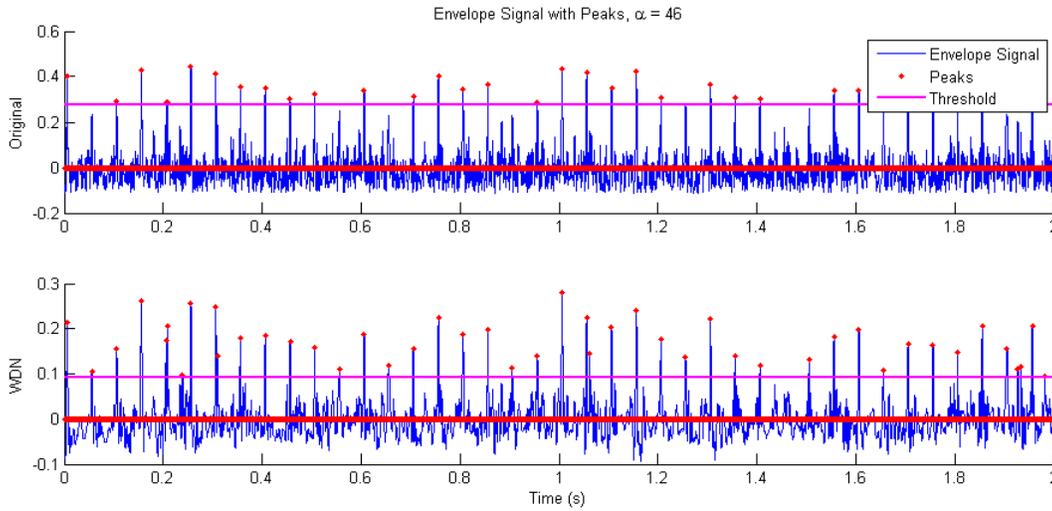
Figure 8. A high threshold causes peaks to be discarded from the first plot, whereas the WDN version of the signal still contains these peaks.

interval plot in Figure 5 shows a "compression" in confidence with variation in threshold scaling α. In the sections that follow, this technique is tested on real-world signals to verify results.

5. REAL SIGNAL TEST RESULTS

5.1. Setup

The Case Western Reserve University bearing data were tested with the WDN algorithm. The precisely seeded faults were created with electro-discharge machining, with the smallest faults at 0.007". A short time-domain waveform is shown of the signal at all stages of the algorithm in Figure 9. The fault frequency is at BPFO (Ball Pass Frequency of Outer raceway).

The db20 wavelet is used to de-noise at 7 levels of decomposi-

tion. The resulting lowpass (scaling filter) cutoff is ≈ 187Hz. With a rotational speed of around 1796 RPM (30Hz), the theoretical maximum fault frequency (BPFI) for the SKF 6205-2RS bearing is approximately 107Hz.

5.2. Frequency-Domain Results

The 0.007" outer raceway fault is distinguishable by the large spectral peak in Figure 10. One small, but noticeable improvement is the BPFO harmonic at ≈ 210Hz, marked with an arrow in the figure, which is removed by WDN.

5.3. Time-Domain Results

The removal of harmonics has implications when attempting to identify faults in the time-domain. Figure 11 shows that the threshold method may pick up harmonics as the domi-
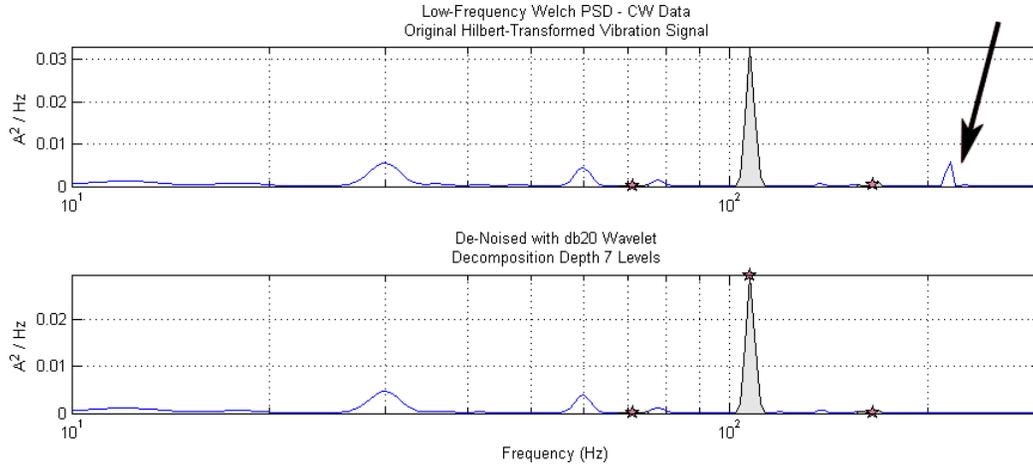
Figure 10. Low frequency PSD of the vibration signal shown in Figure 9. The fault frequency BPFO is approximately 107Hz.
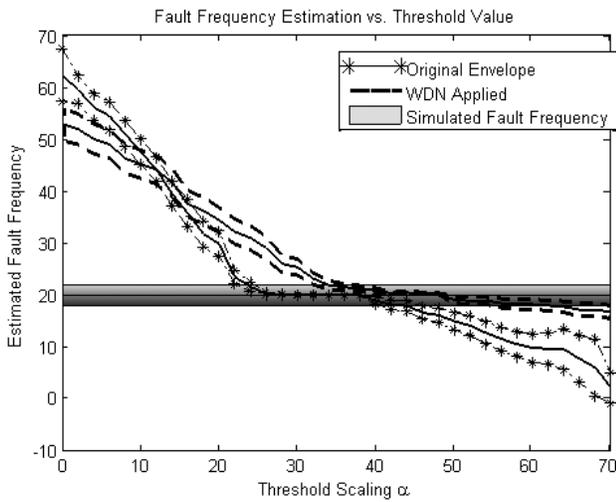
Figure 5. Estimated fault frequency vs. threshold level for the noisy pulse signal shown in 3, periodic at 20Hz. By thresholding the peaks of this envelope at a variable level α, the de-noised envelope signal is shown to more accurately reflect periodicity at the fault frequency.

nant envelope peak frequency. This issue arises when dealing with the BPFO and BSF (Ball Spin Frequency), since their harmonics may be near BPFI. In this case, low α makes estimation inaccurate for the non de-noised signal. With WDN applied, this method works well for low thresholds. At large negative values of α, the algorithm is simply measuring the distance between local maxima.

5.4. Remarks

WDN successfully improved the time-domain fault identification method by reducing its dependence on α. Other data from the Case Western bearing dataset was tested, with similar results.
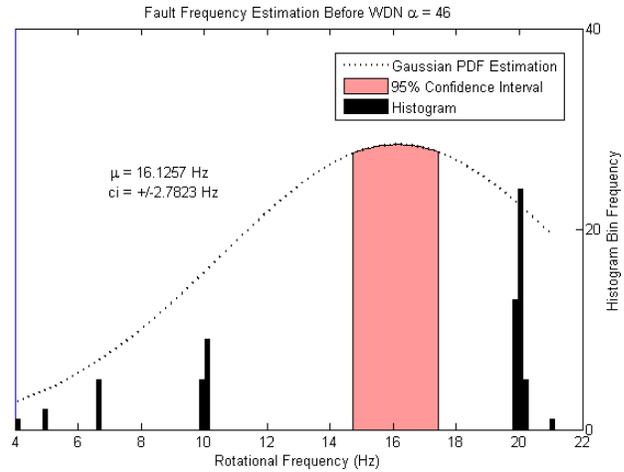
Figure 6. Histogram of impulse frequencies, before WDN. These values are derived from the threshold shown in Figure 8. Low frequency content is a result of the threshold missing lower-amplitude peaks.

6. CONCLUSION

In this paper, we have presented a method to improve detection confidence in fault identification using wavelet denoising. The method deals with the myriad of in-band noise sources in narrowband vibration signals without *a priori* noise statistics.

Decomposition techniques are more suitable to detecting smooth signals, therefore, WDN is applied after envelope demodulation. This yields better results than attempting to de-noise a broadband vibration signal, as in (Qui et al., 2006).

For general purpose signal conditioning, wavelet de-noising is a low-risk, widely applicable technique. Donoho (1995) proves that the gains in noise reduction outweigh the costs of removing low-energy details from the signal. Therefore, unless computational limitations are critical, there is little reason
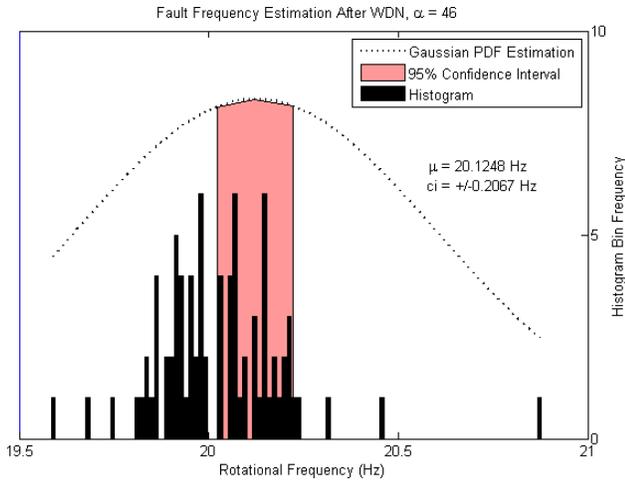
Figure 7. Histogram of impulse frequencies, after WDN. With the same threshold as Figure 6, more peaks are included in the measure, at the proper fault frequency (around 20Hz).

not to utilize such an algorithm.

While this paper demonstrates the function of WDN in the context of demodulated vibration signals, it also serves as a guide for parameter choice. The number of parameters that control this algorithm can be unwieldy, but some sensible decisions and simplifying assumptions allow for ease of use:

- ϕ – wavelet type – In this paper, we decide upon the Daubechies wavelet for its flat passband characteristics. This choice allows for accurate representation of signal proportions in the scale-space domain. In our experiments, the db20 wavelet is sufficient. We choose a high order wavelet, so that the passband cutoff is sharp. This allows for a high number of decompositions without compromising the amplitudes of coefficients in the lower passbands.

- λ – soft threshold level – This value was derived by Donoho (Donoho, 1995) to be a function of noise variance, which is unknown. We simplify this choice by assuming the in-band noise is white, so that the resulting threshold is a function of signal variance, which is known.

- n_d – decomposition level – For our applications, we are searching for energy in the baseband (a demodulated AM-signal). The maximum frequency of a bearing fault is the inner raceway fault frequency (BPFI). Therefore, the decomposition level must not be so high as to place the lowest subband cutoff below this frequency.
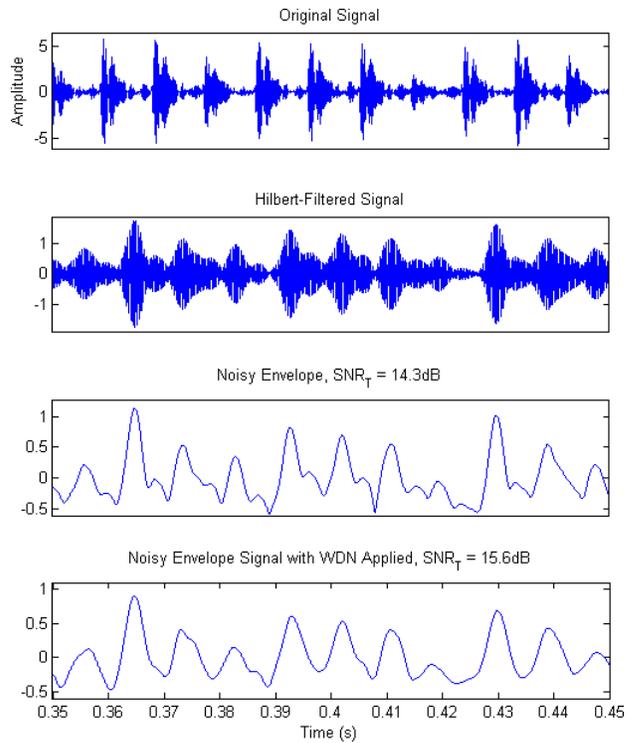
Figure 9. The time-domain waveform of a seeded fault with 0.007" diameter in the outer-raceway at all stages of the algorithm.
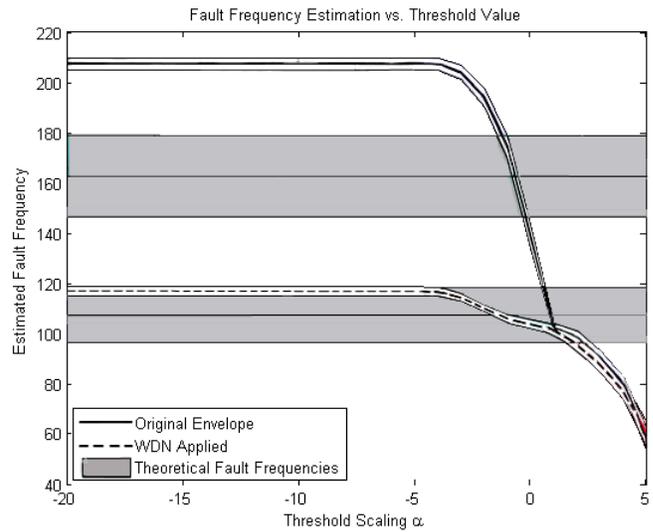
Figure 11. Estimated fault frequency 95% confidence intervals vs. threshold level for the vibration signal shown in 9. This figure demonstrates the importance of removing vibrational harmonics from envelope signals when using a peak thresholding method.

7. FUTURE WORK

To carry this research one step further, it is recommended that power levels be trended over long timescales. The improvements provided by WDN have yet to be tested for evaluation of RUL. It may be hypothesized that, due to the early-detection and confidence improvements demonstrated in this paper, any RUL measure will benefit from earlier, more accurate fault specifics.

With reference to those algorithms tested by Qui et al. (Qui et al., 2006), a direct comparison between wavelet filtering and WDN was never performed, but may be warranted. The WDN algorithm as presented in this paper requires minimal interaction to improve results, where wavelet filtering requires some recursion to tune parameters. Their relative speeds and effectiveness may be a worthwhile measurement.

REFERENCES

Bearing data center. (2013). Case Western Reserve University. Retrieved Jan 2014, from `http://csegroups.case.edu/bearingdatacenter/`

Bozchalooi, I. S., & Liang, M. (2007, July). A combined spectral subtraction and wavelet de-noising method for bearing fault diagnosis. *IEEE Proceedings of the 2007 American Control Conference*, 2533-2538.

Daubechies, I. (1992). *Ten lectures on wavelets*. Philadelphia, PA: Society for Industrial and Applied Mathematics.

Donoho, D. L. (1995). De-noising by soft-thresholding. *IEEE Transactions on Information Theory*, *41*(3), 613-627.

Donoho, D. L., & Johnstone, I. M. (1994). Ideal spatial adaptation via wavelet shrinkage. *Biometrika*, *81*, 425-455.

McFadden, P., & Smith, J. (1984). Model for the vibration produced by a single point defect in a rolling element bearing. *Journal of Sound and Vibration*, *96*, 69-82.

Qui, H., Lee, J., Lin, J., & Yu, G. (2006). Wavelet filter-based weak signature detection method and its application on roller element bearing prognostics. *Journal of Sound and Vibration*, *289*, 1066-1090.

Rioul, O., & Vetterli, M. (1991, October). Wavelets and signal processing. *IEEE Signal Processing Magazine*, *8*(4), 14-38.

Waters, N., & Beaujean, P. (2013). Targeting faulty bearings for an ocean turbine dynamometer. *International Journal of Prognostics and Health Monitoring*, *4*, 1-15.

BIOGRAPHIES

Edward M. Bertot is an engineer and musician from Miami, Florida. He graduated from the University of Miami (Coral Gables, FL) with a B.S. in Music Engineering Technology in 2010, and recently received an M.S. in Electrical Engineering at Florida Atlantic University (Boca Raton, FL). His current research interests include signal processing techniques, with emphasis on audio.

Pierre-Philippe Beaujean received the Ph.D. degree in ocean engineering from Florida Atlantic University in 2001. He is an Associate Professor at the Department of Ocean and Mechanical Engineering, Florida Atlantic University. He specializes in the field of underwater acoustics, signal processing, sonar design, data analysis, machine health monitoring, and vibrations control. Dr. Beaujean is an active member of the Acoustical Society of America (ASA), the Institute of Electrical and Electronic Engineers (IEEE) and the Marine Technology Society (MTS).

David J. Vendittis [Ph.D. (Physics) - American University, 1973] is a Research Professor (part time) at the Department of Ocean and Mechanical Engineering, Florida Atlantic University. Additionally, he is the Technical Advisory Group (ASA/S2) chairman for an International Standards Organization subcommittee, ISO/TC108/SC5 - Machinery Monitoring for Diagnostics. This committee writes international standards that support Machinery Condition Monitoring for Diagnostics. He was appointed to this position by the Acoustical Society of America (ASA).