

Cyclostationary Vibration Analysis and Advanced Signal Pre-Processing for Bearing Fault Detection in Electric Drive Units

Tim Herrmann

Institute of Machine Components, University of Stuttgart, Germany.
E-mail: tim.herrmann@ima.uni-stuttgart.de

Martin Dazer

Institute of Machine Components, University of Stuttgart, Germany.
E-mail: martin.dazer@ima.uni-stuttgart.de

Rolling bearings are critical components for the reliable operation of rotating machinery. As raceway damage is one of the most common failure modes, continuous condition monitoring is essential for its early detection. This enables a predictive maintenance strategy, ensuring cost-effective and optimized repairs. To enhance the model's fidelity to real-world conditions, a complete automotive electric drive unit, including its various excitations is analyzed using a multibody simulation. Since the cyclostationary nature of raceway faults prevents unambiguous detection using conventional frequency analysis, the envelope spectrum is the established diagnostic standard. However, its effectiveness is severely limited in integrated electric drives. Due to the high-energy deterministic background noise generated by the transmission, the weak stochastic fault signatures are frequently masked, preventing reliable detection even in the envelope domain. To overcome this limitation, this paper evaluates advanced signal pre-processing techniques designed to enhance the signal-to-noise ratio prior to demodulation. Among the investigated methods, cepstrum pre-whitening (CPW) has proven particularly effective in suppressing deterministic excitations, thereby removing superimposed gear mesh excitations from the signal. This significantly improves early fault detection capability. When combined with minimum entropy deconvolution (MED), the impulsiveness of the signal is further optimized, allowing the damage characteristics to be highlighted even more distinctly.

Keywords: Raceway damage, envelope spectrum, cyclostationary signal, cepstrum pre-whitening, minimum entropy deconvolution.

1. Introduction

The electrification of the automotive powertrain is a key driver for reducing emissions and achieving sustainable mobility. Modern electric vehicles increasingly rely on highly integrated Electric Drive Units (EDUs), which combine the electric machine, transmission, and power electronics into a single compact housing. While this integration offers significant advantages in terms of power density and weight, it introduces new challenges for reliability and maintenance (Tiboni et al., 2022).

Rolling element bearings are among the most critical components within these drive units. Due to the high rotational speeds and variable load conditions characteristic of electric drives, bearings are susceptible to fatigue and wear. Raceway damage, in particular, represents a frequent failure mode that can lead to catastrophic system break-

downs if not detected at an early stage. Consequently, continuous condition monitoring is essential to enable predictive maintenance strategies, thereby minimizing downtime and repair costs.

Vibration analysis is the standard method for monitoring rotating machinery. However, detecting bearing faults in an automotive EDU is complex. The vibration signal is dominated by high-energy deterministic excitations from the gear mesh, which often mask the much weaker, impulsive signatures of early bearing damage. Furthermore, unlike the deterministic signals generated by gears, bearing fault signatures caused by raceway damage are stochastic in nature. They exhibit cyclostationary behavior, a statistical property where the signal's energy varies periodically over time but spreads across a wide frequency band (Borghesani et al., 2022). Standard

frequency spectrum analysis is therefore insufficient for unambiguous fault detection.

To address this challenge, the envelope spectrum is widely established as a diagnostic tool (Franke, 2022). Yet, its effectiveness depends heavily on the quality of the input signal. Various signal pre-processing techniques have been proposed to suppress deterministic noise and enhance the impulsive fault features. However, validating these methods on real-world systems is difficult due to the lack of run-to-failure data with precisely known ground truth.

This paper presents a comparative study of signal processing techniques for the early detection of raceway damage in EDUs. To overcome the limitations of experimental data availability, we employ a high-fidelity Multibody Simulation (MBS) of a complete automotive electric drive. This simulation allows for the generation of synthetic vibration data with controllable fault severities and operating conditions. We specifically evaluate the effectiveness of Cepstrum Pre-Whitening (CPW) and Minimum Entropy Deconvolution (MED) as pre-processing steps for envelope analysis, benchmarking them against established separation techniques. The results demonstrate that distinguishing between deterministic and cyclostationary content is crucial for reliable raceway damage diagnosis in complex drivetrains.

2. System Modeling and Simulation

To evaluate the diagnostic performance of signal processing algorithms under reproducible conditions, we use a high-fidelity Multibody Simulation (MBS) of a generic automotive electric drive system. This approach allows for the generation of vibration data with precise fault labels, circumventing the limitations of experimental setups where ground truth regarding the exact fault severity is often unavailable.

2.1. Electric drive unit configuration

The system under investigation represents a typical integrated Electric Drive Unit (EDU) used in battery electric vehicles. The drivetrain consists of a high-speed electric traction motor, a two-stage helical gear transmission, and a differential. The

structural dynamics of the housing are included to realistically capture the transfer path of structure-borne noise from the internal components to the sensor locations.

The study focuses on the deep-groove ball bearings supporting the transmission input shaft. These components are subjected to the highest rotational speeds and highest stresses within the system, making them critical for reliability. To emulate a standard condition monitoring setup, two virtual accelerometers are positioned on the housing surface, capturing vibration in three translational degrees of freedom (x, y, z). To ensure realistic signal transfer behavior, the virtual sensor positions are based on a configuration validated against the physical drive unit. The signal is sampled at a standard industrial rate suitable for envelope analysis.

2.2. Stochastic fault modeling

A key aspect of this study is the realistic modeling of raceway faults. While simplified theoretical models assume a strictly periodic impulse train $T_k = k \cdot T$, real-world bearing faults exhibit cyclostationary behavior due to the complex kinematics of the rolling elements (Antoni and Randall, 2002). To capture the true stochastic nature of the signal, we implement a Mixed Cyclostationary Model. Following the theoretical framework established by Borghesani et al. (2022), the arrival time T_k of the k -th impact is defined as the superposition of two stochastic processes:

$$T_k = t_k + \tau_k, \quad (1)$$

where t_k represents the cumulative effect of cage slip and τ_k accounts for the roller jitter. These components capture two distinct kinematic phenomena:

- (1) **Cage Slip:** The speed of the cage fluctuates due to slip, preventing the time between impacts from being strictly constant. This is modeled as a random walk process, where the mean impact time t_k depends on the previous impact t_{k-1} :

$$t_k = t_{k-1} + \Delta t_k \quad (2)$$

$$\text{with } \Delta t_k \sim \mathcal{N}(\mu_\Delta, \sigma_\Delta^2). \quad (3)$$

Here, μ_Δ corresponds to the nominal fault period, while the variance σ_Δ^2 defines the intensity of the cumulative timing deviation.

- (2) **Roller Jitter:** The rolling elements possess a degree of freedom due to clearance within the cage. Their impact position varies stochastically around the mean cage position. This is modeled as a zero-mean Gaussian deviation τ_k added independently to each impact:

$$\tau_k \sim \mathcal{N}(0, \sigma_\tau^2). \quad (4)$$

The fault impacts are generated as discrete force excitations acting radially between the inner and outer raceways. Physically, the passage of a rolling element over a localized surface defect (pitting) results in a sharp collision. This impact phenomenon is mathematically idealized as a Dirac delta function (Borghesani et al., 2022). Based on the calculated stochastic timing sequence T_k an impulsive force vector is applied to the respective rigid bodies in the MBS environment. To ensure sufficient temporal resolution of these short-duration impact dynamics, the internal simulation runs at a high oversampling rate before being downsampled to the target sensor frequency.

2.3. Data generation and operating points

The dataset covers stationary operating points at various vehicle speeds representing typical urban and highway driving profiles. For each operating condition, a vibration signal with a duration of 10.0 seconds was simulated and subsequently divided into ten 1.0 second segments for vibration analysis.

To quantify the severity of the defect, the Damage Intensity Factor is introduced. This normalized metric scales the magnitude of the fault impact force relative to the instantaneous bearing load, ensuring that damage levels remain comparable across different operating points and load cases. The simulated range covers the critical transition from early to advanced faults. Based on the signal-to-noise characteristics, a Damage Intensity Factor greater than 0.2 results in a very dominant damage signature. In contrast, a Damage Intensity Factor below 0.1 represents an early fault stage where the impulsive features are heav-

ily masked by the background vibration and structural resonances.

The analysis includes faults on both the inner and outer raceways. This variation generates distinct kinematic fault frequencies, ensuring that the diagnostic methods are evaluated across different signal signatures within the same mechanical system.

3. Signal Processing Framework

The detection of early bearing faults in complex drive trains requires separating the weak, cyclostationary fault signatures from the high-energy, deterministic background noise generated by the transmission. To achieve this, we implement a multi-stage signal processing pipeline as seen in Figure 1.

The pipeline processes the raw Acceleration Signal through Deterministic Signal Suppression Methods to remove high-energy periodic masking. Subsequently, Deconvolution Methods for Impulse Enhancement are applied to recover and sharpen the fault signature. Finally, the Squared Envelope Spectrum (SES) is computed to enable Peak Detection at Bearing Orders.

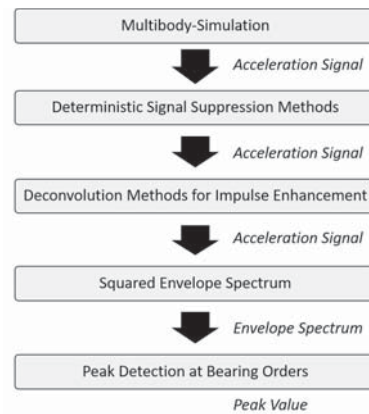


Fig. 1. Schematic overview of the signal processing pipeline combining Cepstrum Pre-Whitening (CPW) and Minimum Entropy Deconvolution (MED).

3.1. Deterministic signal suppression methods

The suppression of high-energy deterministic noise is the first critical step in the pipeline. There are two fundamental approaches: residual-based techniques, which attempt to reconstruct and subtract the deterministic component, and spectral whitening, which equalizes the signal energy across the frequency band.

3.1.1. Residual-based techniques

Time Synchronous Averaging (TSA) is widely known for its computational simplicity and established status as the standard technique for gear vibration analysis. It relies on the assumption of strict periodicity synchronous with the shaft. The vibration signal is averaged over multiple shaft revolutions to isolate the synchronous content. Subtracting this average from the raw signal yields a residual that is theoretically free of deterministic gear mesh excitations (Randall et al., 2011).

Discrete Random Separation (DRS), in contrast, operates in the frequency domain. This method exploits the spectral distinction between signal types: it suppresses discrete frequency lines with high signal-to-noise ratios, effectively acting as a self-adaptive filter to separate deterministic harmonics from the broadband stochastic background noise (Randall and Antoni, 2011). These methods serve as a baseline to demonstrate the specific challenges imposed by the cyclostationary nature of bearing faults in the presence of slip.

3.1.2. Cepstrum pre-whitening

Cepstrum Pre-Whitening (CPW) is employed to suppress the dominant deterministic components that mask the broadband bearing fault signature. In the frequency domain, these deterministic components manifest as distinct peaks, masking the broadband energy distribution of bearing faults. To address this, the concept of the real cepstrum $C_{real}(\tau)$ is utilized (Peeters et al., 2017), which is defined as the inverse Fourier transform of the log-magnitude spectrum:

$$C_{real}(\tau) = \mathcal{F}^{-1} \{ \ln |\mathcal{F}\{x(t)\}| \} \quad (5)$$

In the cepstral domain, families of periodic harmonics (e.g., gear mesh series) collapse into distinct peaks known as *rahmonics*. Theoretically, removing these deterministic components corresponds to setting the non-zero quefrencies in the cepstrum to zero (liftering) (Peeters et al., 2017). Mathematically, however, this operation is equivalent to flattening the magnitude spectrum to unity while strictly preserving the phase information. This allows for a computationally efficient implementation directly in the frequency domain:

$$x_{cpw}(t) = \mathcal{F}^{-1} \left\{ \frac{X(f)}{|X(f)|} \right\} \quad (6)$$

$$= \mathcal{F}^{-1} \left\{ 1 \cdot e^{j\phi(f)} \right\} \quad (7)$$

where $X(f)$ is the Fourier transform of the input signal and $\phi(f)$ represents its phase. The crucial advantage of this "whitening" process is that the impulsive signature of the bearing fault is encoded primarily in the phase alignment of broadband frequencies, whereas the gear noise relies on high spectral amplitudes. By normalizing the amplitude but retaining the phase, the dominant gear signals are suppressed, while the fault impulses are preserved and exposed from the background noise as can be seen in Figure 2.

3.2. Deconvolution methods for impulse enhancement

While the preceding suppression stage effectively removes the dominant deterministic signal components, it does not correct the physical distortion of the fault impulse caused by the transmis-

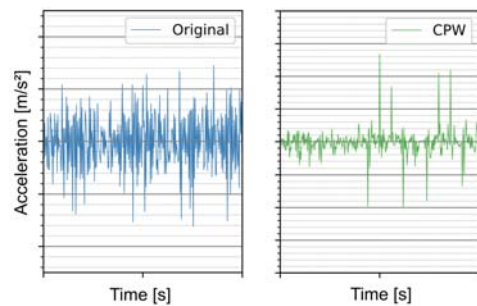


Fig. 2. Qualitative comparison: Raw acceleration signal (dominated by gears) vs. CPW-processed signal (revealing impacts).

sion path. As the sharp impact force propagates through the complex housing structure, it undergoes dispersion and attenuation. This effectively smears the impulse in the time domain, reducing its peak amplitude and kurtosis. To counteract this structural transfer function and recover the sharpness of the source signal, deconvolution is applied as the second major stage of the processing pipeline.

3.2.1. Minimum entropy deconvolution

Minimum Entropy Deconvolution (MED) is an inverse filtering technique designed to counteract this effect (Wiggins, 1978). Specifically, this study employs the "Minimum Entropy Deconvolution Adjusted" (MEDA) algorithm (McDonald and Zhao, 2017), which is optimized for the continuous nature of machinery vibration signals compared to the original seismic model. This optimized variant is referred to as MED throughout this paper.

It searches for an optimal Finite Impulse Response (FIR) filter f that recovers the impulsive source signal $y(n)$ from the measured signal $x(n)$:

$$y(n) = \sum_{l=0}^{L-1} f(l)x(n-l) \quad (8)$$

The filter coefficients are determined iteratively by maximizing the Kurtosis (normalized fourth central moment), which serves as a measure of the signal's "spikiness". For this purpose, the objective function $O_k(f)$ is defined as the normalized fourth moment of the filtered signal:

$$\text{Maximize } O_k(f) = \frac{\sum y^4}{(\sum y^2)^2} \quad (9)$$

By maximizing this ratio, MED emphasizes the impulsive components associated with bearing impacts. The optimal filter coefficients are found when the gradient of the Kurtosis with respect to the filter f is zero. However, MED is prone to locking onto single, spurious impulses if not constrained. Therefore, applying CPW prior to MED is strategically important: CPW removes the dominant periodic gear signals, ensuring that the MED filter optimizes specifically for the repetitive

bearing fault impulses rather than gear mesh artifacts.

3.2.2. Selection of deconvolution strategy

Prior to establishing MED as the core filter, a comparative evaluation of alternative deconvolution techniques, including OMEDA, MCKD, MOMEDA, MOMEDAS, and ACY-CBD, was conducted (McDonald et al., 2012; McDonald and Zhao, 2017; Wang et al., 2022).

Target-specific methods (MCKD, MOMEDA) were discarded as they require the fault period as an a-priori input, necessitating computationally prohibitive multi-pass processing for every potential fault order. Furthermore, MOMEDA demonstrated a critical tendency to generate false positives by enforcing periodicity even on noise signals where no fault was present.

Blind estimators (MOMEDAS, ACYCBD) and OMEDA, as an analytical solution, were ruled out due to excessive execution times exceeding numerical MED by factors up to 2000 without yielding superior diagnostic clarity. Notably, the analytical OMEDA approach frequently failed to recover the periodic fault train, instead converging on single, spurious impulses. In contrast, the iterative MED algorithm demonstrated superior stability, consistently enhancing the repetitive fault signatures without locking onto isolated artifacts.

Consequently, the iterative MED algorithm was identified as the optimal trade-off between blind impulse recovery, computational efficiency, and robustness against false alarms.

3.3. Envelope analysis and peak detection

The final stage of the diagnostic pipeline involves the demodulation of the enhanced signal and the extraction of the quantitative fault indicators. First, the Squared Envelope Spectrum (SES) is generated. The signal processed by the pipeline is demodulated using the Hilbert transform to extract the signal envelope. The Fourier transform of this squared envelope reveals the repetition frequency of the impacts (Franke, 2022). The application of deterministic suppression and impulse sharpening ensures that the demodulated signal is dominated by the fault signature. As a result, the

SES reveals clear, high-contrast spectral peaks, effectively overcoming the limitations of standard envelope analysis.

To conclude the analysis, a peak detection algorithm is applied to the SES. The algorithm searches for spectral amplitudes at the characteristic kinematic fault orders, specifically the Ball Pass Frequency Outer (BPFO) and Inner (BPFI). Since the exact fault frequency fluctuates slightly due to the stochastic cage slip described in Section 2.2, the search is performed within a defined tolerance window around the theoretical kinematic orders. The maximum amplitude identified within this window serves as the final diagnostic feature ("Peak Value"). Crucially, unlike amplitudes in conventional frequency spectra, this metric quantifies the distinctness of the cyclic signature against the background noise. Therefore, it reflects the statistical confidence of the detection rather than the physical severity of the damage.

4. Results

The comparative analysis evaluates the performance of the investigated signal processing techniques regarding the key requirements for a reliable on-board monitoring system. The analysis follows a logical progression: it begins with a system observability analysis to validate the measurement configuration and signal transfer behavior. Building on this baseline, the fundamental capability to differentiate between healthy and damaged states is assessed to establish a robust detection threshold. Subsequently, the evaluation focuses on the sensitivity to early faults, benchmarking the methods' performance in the low damage intensity range. Finally, the robustness against environmental factors, specifically sensor noise, is examined to ensure applicability under realistic operating conditions.

4.1. System observability analysis

Prior to the detailed benchmarking, the influence of sensor placement and fault location was briefly assessed. The simulation results confirm that fault signatures are principally observable in all three translational directions (x, y, z). Ultimately, the clarity of the fault signatures is determined by the

specific interaction between the bearing's location and the sensor position on the housing structure. To reflect a realistic application scenario, the sensors are not placed directly at the fault source but on the external housing surface, introducing a representative transfer path with natural signal damping. Despite these geometric variations, the fault patterns remained consistently detectable.

Regarding operating conditions, the analysis revealed that urban driving profiles yielded superior diagnostic clarity compared to high-speed highway scenarios. To ensure a consistent baseline for the following comparison, the analysis focuses on the sensor channel yielding the highest energy content for each respective fault scenario within this optimal operating range.

4.2. Differentiation of healthy and damaged states

To establish a reliable detection threshold preventing false alarms, a statistical analysis of the non-damaged orders was conducted. Based on $n = 2560$ healthy data points across all operating conditions, Figure 3 illustrates the distribution of peak amplitudes in the envelope spectrum where no fault is present. The results show that for all methods, the 90th percentile of the noise peaks remains below an amplitude of 0.15. The 99th percentile is consistently located below 0.3, with the vast majority of outliers remaining below 0.25. Consequently, a conservative detection threshold of $A_{th} = 0.25$ was defined. Any spectral peak exceeding this value can be classified as a fault with a high degree of confidence, effectively minimizing the risk of false positives.

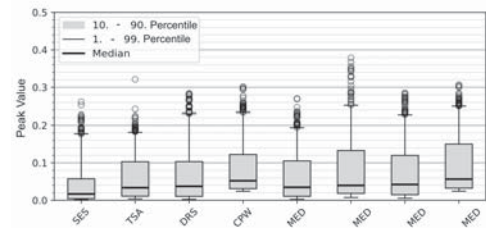


Fig. 3. Statistical distribution of peak amplitudes at non-fault orders ($n = 2560$) to determine the baseline noise floor.

4.3. Early detection capability

The primary objective of a predictive maintenance system is the identification of faults at the earliest possible stage. Figure 4 illustrates the mean detection values ($n = 10$) for each method plotted against the fault intensity.

The analysis reveals that CPW distinguishes itself significantly from other methods for detection of early fault intensities. Even for the lowest simulated damage levels, CPW yields feature amplitudes well above the detection threshold, confirming its ability to effectively suppress deterministic gear mesh excitations. While the SES consistently yields the poorest performance due to the masking effect of the transmission, the behavior of the residual-based methods shows distinct differences. TSA alone provides only marginal improvements over the SES. In contrast, Discrete Random Separation (DRS) noticeably outperforms TSA, suggesting that its frequency-domain filtering is inherently more robust against the time-domain jitter induced by cage slip. However, even DRS fails to match the sensitivity of CPW for early fault intensities.

When combined with MED, all suppression methods show improved performance. TSA-MED achieves a substantial increase in detectability, ranking third overall. Nevertheless, the combination of CPW and MED proves to be the superior approach. The suboptimal performance of TSA and DRS compared to CPW indicates that

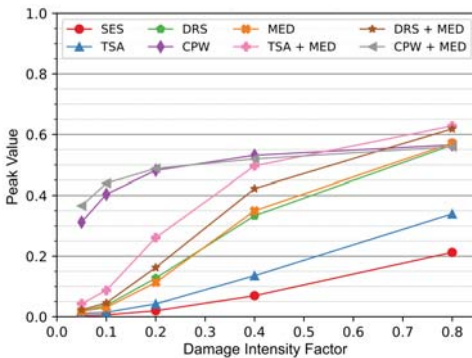


Fig. 4. Comparison of the early detection capability across increasing fault intensities (Bearing X).

strictly separating deterministic lines is less effective than global spectral equalization. Even after removing synchronous components, the residual signal contains significant background noise. The CPW-MED pipeline overcomes this by combining spectral whitening with blind deconvolution. By removing the deterministic dominance first, CPW ensures that the MED optimization maximizes the kurtosis of the fault impulses rather than locking onto gear mesh residuals, leading to the highest observed sensitivity.

4.4. Robustness against sensor noise

Real-world automotive environments are characterized by significant background noise. To evaluate robustness, the acceleration signals were corrupted with additive white Gaussian noise at Signal-to-Noise Ratios (SNR) ranging from 40 dB (high quality) down to 5 dB (highly noisy). Figure 5 shows the performance of the full processing pipelines (DRS-MED vs. CPW-MED) under these degrading conditions.

The analysis reveals that under realistic noise conditions (down to 20 dB), both pipelines exhibit stable performance. However, a critical divergence is observed in the low-SNR range (10 dB and below). The DRS-MED pipeline shows a significant sensitivity to noise, particularly for early faults (Damage Intensity Factor < 0.2), where the detection capability collapses almost completely at 10 dB.

In contrast, the CPW-MED pipeline demonstrates exceptional resilience. Even at an SNR of 10 dB, it maintains high peak values for early faults, significantly outperforming the DRS-based approach. Since the deconvolution stage (MED) is identical in both pipelines, this superior robustness can be directly attributed to the CPW algorithm. While DRS relies on the separation of discrete lines, CPW's spectral whitening proves to be the decisive factor in maintaining signal integrity. Consequently, the CPW-MED pipeline ensures reliable monitoring down to an SNR of 10-20 dB, covering the realistic range of sensor noise expected in an electric drive unit.

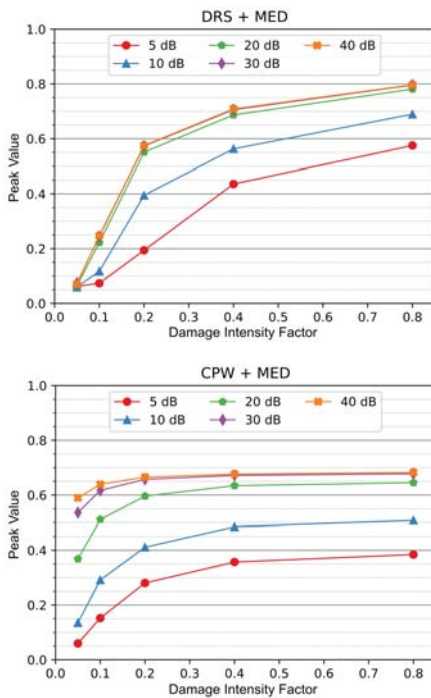


Fig. 5. Robustness evaluation of DRS-MED and CPW-MED regarding additive sensor noise (Bearing Y).

5. Conclusion

This study presented a comprehensive evaluation of signal processing techniques for the condition monitoring of automotive electric drive units. By utilizing a high-fidelity MBS, it was possible to generate a validated dataset with precise ground truth, covering the critical transition from early to advanced fault stages. The analysis highlighted that the primary diagnostic challenge in integrated EDUs is the high-energy masking effect of the gear mesh, which obscures the weak, cyclostationary signatures of raceway faults.

The comparative analysis identifies the combination of CPW and MED as the optimal methodology, consistently outperforming residual-based methods like TSA and DRS. The study reveals that these established techniques are limited either by signal smearing due to the cyclostationary nature of the fault signatures (TSA) or by sensitivity to colored background noise (DRS). In contrast,

the spectral whitening approach of CPW proves inherently more robust, effectively isolating the fault signature from both kinematic fluctuations and system-inherent noise. Consequently, the proposed pipeline demonstrates exceptional stability, maintaining detection capability for early faults even at low signal-to-noise ratios.

References

- Antoni, J. and R. Randall (2002). Differential diagnosis of gear and bearing faults. *Journal of Vibration and Acoustics* 124(2), 165–171.
- Borghesani, P., W. Smith, R. Randall, J. Antoni, M. El Badaoui, and Z. Peng (2022). Bearing signal models and their effect on bearing diagnostics. *Mechanical Systems and Signal Processing* 174.
- Franke, D. (2022). *Wälzlagerdiagnose an Maschinensätzen* (1 ed.). Springer Vieweg Berlin, Heidelberg.
- McDonald, G. L. and Q. Zhao (2017). Multipoint optimal minimum entropy deconvolution and convolution fix: Application to vibration fault detection. *Mechanical Systems and Signal Processing* 82, 461–477.
- McDonald, G. L., Q. Zhao, and M. J. Zuo (2012). Maximum correlated kurtosis deconvolution and application on gear tooth chip fault detection. *Mechanical Systems and Signal Processing* 33, 237–255.
- Peeters, C., P. Guillaume, and J. Helsen (2017). A comparison of cepstral editing methods as signal pre-processing techniques for vibration-based bearing fault detection. *Mechanical Systems and Signal Processing* 91, 354–381.
- Randall, R., N. Sawalhi, and M. Coats (2011, 06). A comparison of methods for separation of deterministic and random signals. *International Journal of Condition Monitoring* 1, 11–19.
- Randall, R. B. and J. Antoni (2011). Rolling element bearing diagnostics—a tutorial. *Mechanical Systems and Signal Processing* 25(2), 485–520.
- Tiboni, M., C. Remino, R. Bussola, and C. Amici (2022). A review on vibration-based condition monitoring of rotating machinery. *Applied Sciences* 12(3), 972.
- Wang, Z., J. Zhou, W. Du, Y. Lei, and J. Wang (2022). Bearing fault diagnosis method based on adaptive maximum cyclostationarity blind deconvolution. *Mechanical Systems and Signal Processing* 162, 108018.
- Wiggins, R. A. (1978). Minimum entropy deconvolution. *Geophysical Research Letters* 5, 21–35.