Cluster-Based Permutation Tests for Automated Identification of Critical Frequency Bands in Vibration Data

Neil Eklund[1], Adam Wechsler[2], Colin Dingley[2], Sherwood Polter[2], Khai Van[2]

[1]*Noblis MSD, Philadelphia, PA, 19112, USA*
neil.eklund@noblismsd.com, neil.eklund@gmail.com, or neil.h.eklund.ctr@us.navy.mil

[2]*Naval Surface Warfare Center Philadelphia Division, Philadelphia, PA, 19112, USA*
adam.s.wechsler.civ@us.navy.mil, colin.j.dingley.civ@us.navy.mil, sherwood.v.polter.civ@us.navy.mil, khai.t.van.civ@us.navy.mil

Abstract

Condition-based maintenance plus for naval applications requires robust embedded health management systems. However, extracting reliable diagnostic features from reciprocating machinery in harsh maritime environments is hindered by severe acoustic interference and non-stationary dynamics. Traditional point-wise statistical methods, such as analysis of variance, suffer from the multiple comparisons problem and select redundant, highly correlated frequency bins. Conversely, unsupervised dimensionality reduction techniques like principle component analysis are susceptible to broadband noise and lack physical interpretability.

To address these limitations, this paper presents a hierarchical spatial-spectral feature selection framework based on cluster-based permutation tests (CBPT). By evaluating multivariate spectral adjacency and incorporating stratified block bootstrap stability selection, the method isolates deterministic kinematic resonances. This non-parametric approach rejects stochastic background noise and mitigates the large-sample paradox without requiring manual frequency band selection.

The framework was validated using multi-channel vibration data from a reciprocating electromechanical testbed subjected to working fluid leakage, intake restriction, and thermal management degradation. To assess generalization, models were evaluated using a chronological train-test split with progressive levels of additive white Gaussian noise added to the test sets. The CBPT-derived feature space maintains a macro F1-score above 0.90 down to a 5 dB signal-to-noise ratio, outperforming standard univariate and broadband extraction baselines under acoustic interference. This methodology yields an independent, low-dimensional feature set suitable for autonomous, embedded prognostic architectures.

1. Introduction

Condition-based maintenance plus (CBM+) advances traditional maintenance paradigms by integrating prognostics and health management (PHM), data analytics, and knowledge-based systems to forecast rather than detect equipment failures. Whereas standard condition-based maintenance (CBM) triggers maintenance actions from real-time asset monitoring, CBM+ uses predictive algorithms, machine learning (ML), and artificial intelligence (AI) to estimate remaining useful life (RUL). This transition from reactive to proactive maintenance optimizes resource allocation, reduces unscheduled downtime, and enhances system reliability (DoD, 2018).

The U.S. Navy advances CBM+ implementation through programs such as the Integrated Condition Assessment System (ICAS). However, fleet deployment presents distinct technical challenges, particularly restricted data transfer between vessels at sea and shoreside analysis centers. The Naval Surface Warfare Center, Philadelphia Division (NSWCPD) is developing embedded automated health management systems (HMSs) within new hull, mechanical, and electrical (HM&E) monitoring and control architectures. Modern ship construction enables the integration of sensor data collection, onboard processing, and predictive analytics directly into the platform.

A major barrier to HMS deployment is the scarcity of representative fault data for training and validating ML models. Consequently, NSWCPD conducts controlled fault testing on critical mechanical systems to generate high-fidelity datasets for model development.

This paper presents a feature selection framework for embedded HMS to identify critical frequency bands in high-dimensional vibration spectra. Due to shipboard mechanical complexity, manual frequency band selection is subjective and prone to missing subtle spectral shifts. To address this limitation, we apply cluster-based permutation tests (CBPT), a non-parametric statistical method, to isolate frequency clusters exhibiting significant deviations across fault conditions. Using fault data from NSWCPD controlled testing, we demonstrate that CBPT reduces feature dimensionality while preserving diagnostic sensitivity. This approach improves onboard predictive model precision and decreases computational demand on embedded architectures, supporting autonomous, real-time health assessment.

1.1. Reciprocating Machinery and Shipboard Acoustic Environment

The system under study is a multi-stage, reciprocating, high-energy fluid power source; two to four units (or more on larger vessels) are installed on nearly all major U.S. Navy surface ships and every submarine in the fleet. Vibration monitoring provides an effective diagnostic approach for this system because mechanical anomalies manifest as spectral shifts before detection by conventional metrics such as temperature or pressure (Fang et al., 2024). Structural deterioration, including valve failure or piston ring thinning, induces distinct signatures detectable by machinery-mounted accelerometers.

Power spectral density (PSD) is a standard feature extraction method in PHM. PSD characterizes signal power distribution across frequencies, transforming complex time-domain vibration signals into frequency-domain profiles. By calculating the mean square value of the vibration amplitude per unit of frequency, PSD yields a normalized measure that remains consistent regardless of the frequency resolution or sampling record length. This statistical method is well suited for analyzing non-periodic or random vibration signals because it isolates energy intensity at spectral peaks that standard Fourier transforms may obscure. Comparing the measured power spectrum against a healthy baseline enables the PHM system to detect early wear and quantify fault severity, supporting data-driven maintenance scheduling and preventing catastrophic equipment failure.

However, the deployed system and its operational environment present a challenging acoustic regime for condition monitoring due to periodic impulsive loading cycles and complex vibration profiles. Unlike the stable signatures of centrifugal systems, this machinery generates high vibration through mechanical impacts, reciprocating inertial forces, and high-energy pressure-discharge pulses. These broadband excitations create an elevated noise floor in the shipboard environment, where dominant energy is non-stationary and masked by auxiliary hull-borne noise (Falzone & Kolodziej, 2017). The resulting low signal-to-noise ratio (SNR) complicates the isolation of discrete frequency components associated with incipient degradation, an issue exacerbated by fluctuating operational loads and variable maritime conditions (Soualhi et al., 2025).

Estimating critical fault frequencies *a priori* is difficult due to interactions between kinematic expectations and operational variables. Although fundamental rotational speeds are theoretically fixed, load-dependent slip, thermal expansion of internal clearances, and non-linear stiffness of mounting interfaces alter the spectral response. Harmonic overlap across mechanical stages often causes spectral smearing, which prevents the resolution of discrete bearing defect frequencies or transient valve-seating events (Trout & Kolodziej, 2016). This uncertainty necessitates an empirical approach to frequency identification, as theoretical kinematic models often fail to capture integrated structural resonances and environmental variations that alter characteristic vibration signatures in naval applications (Lee et al., 2017; Jardine et al., 2006).

2. CLUSTER-BASED PERMUTATION TESTS

Identifying anomalous spectral components in high-resolution vibration data presents a difficult statistical challenge. In environments where signals are non-stationary and influenced by sea states, engine loads, and acoustic interference from nearby equipment, standard frequentist approaches fail to balance sensitivity with statistical rigor. The CBPT provides a non-parametric alternative designed for high-dimensional, dependent data structures by exploiting the correlation between adjacent frequency bins.

2.1. Limitations of Point-wise Inference

The primary hurdle in spectral analysis is the multiple comparisons problem (MCP). When assessing vibration data across a broad spectrum (e.g., 0–10 kHz at 1 Hz resolution), performing a separate *t*-test at each of the 10,001 frequency bins inflates the Type I error rate. At a standard significance level of $\alpha=0.05$, approximately 500 false positives occur purely by chance, making it nearly impossible to distinguish true mechanical faults from stochastic noise. Standard solutions such as the Bonferroni correction control the familywise error rate (FWER) by dividing α by the number of comparisons, but this approach is notoriously conservative. In high-resolution spectral data, it reduces the significance threshold to such an extreme that subtle, broadband shifts, such as early-stage bearing wear or minor sideband development, are frequently masked. Similarly, false discovery rate (FDR) corrections like the Benjamini-Hochberg procedure (Benjamini & Hochberg, 1995) treat each frequency bin independently, ignoring spectral continuity and physical signal structure.

In vibration spectra, adjacent frequency bins are interdependent due to broadband mechanical resonances and digital signal processing (DSP) phenomena, particularly spectral leakage from window functions. Furthermore, standard PSD estimation using overlapping time segments (e.g., Welch's method) introduces mathematical covariance across neighboring bins. This spectral correlation violates the independence assumption required by point-wise statistical tests. CBPT exploits this dependency, treating bin-to-bin correlation as an informative property rather than a statistical nuisance.

2.2. Conceptual Framework: The "Cluster"

Instead of evaluating isolated frequency bins, CBPT analyzes continuous spectral regions by clustering adjacent bins. The procedure first identifies all frequency bins where an uncorrected test statistic exceeds a preliminary threshold, typically $p < 0.05$. Spatially adjacent bins meeting this criterion are grouped into "clusters". To quantify spectral deviation, an independent two-sample Welch's *t*-statistic is computed for each frequency bin *f*:

$$t(f) = \frac{\bar{X}_d(f) - \bar{X}_h(f)}{\sqrt{\frac{s_d^{2(f)}}{N_d} + \frac{s_h^{2(f)}}{N_h}}} \quad (1)$$

Where $\bar{X}_d(f)$ and $\bar{X}_h(f)$ represent the mean log-transformed PSD amplitudes, expressed in decibels (dB), for the damaged and healthy conditions at frequency *f*. Variables $s_d^2(f)$ and $s_h^2(f)$ denote the corresponding sample variances of these log-transformed amplitudes, while N_d and N_h represent the sample sizes for each condition. The raw PSD estimates are log-transformed prior to calculating the *t*-statistic to stabilize spectral variance. Within this non-parametric framework, the normality assumption is not required for final statistical inference; Welch's *t*-statistic functions solely as a cluster-forming metric to establish spectral adjacency. Statistical significance and empirical p-values are derived from the distribution-free permutation distribution formulated in equation (4), below.

A cluster C_k is formally defined as a contiguous set of frequency bins where the absolute point-wise statistic exceeds a predefined critical threshold, $|t(f)| > t_{crit}$. The cluster-level statistic S_k is then computed as the discrete sum (or integral) of the *t*-values within that spectral envelope:

$$S_k = \sum_{f \in C_k} t(f) \quad (2)$$

This formulation prioritizes spectrally broad, high-magnitude deviations, enabling the detection of low-amplitude spectral shifts that fail point-wise significance tests. Aggregating *t*-values across contiguous bins captures spectral coherence across the frequency band.

2.3. The Permutation Statistic

Fisher (1935) established the foundation of permutation-based resampling to determine statistical significance. However, exact permutation tests remained computationally intractable for decades. For example, calculating Fisher's Exact Test value for even a modest dataset of two groups with 10 observations each requires evaluating 184,756 (20 choose 10) possible combinations‱a calculation that was impractical before modern computing. The scientific adoption of resampling methods accelerated after Efron (1979) formalized the bootstrap, demonstrating that computational resampling could replace asymptotic approximations.

Diaconis and Efron (1983) broadened the adoption of these concepts, establishing the foundation for computer-intensive statistical methods. By demonstrating that computational power could replace complex analytical formulas and that resampling could estimate empirical sampling distributions, these works established the basis for non-parametric procedures in modern spectral analysis.

CBPT extends this methodology by using label exchangeability to construct empirical null distributions tailored to specific spectral noise profiles. In this approach, cluster significance is determined through permutation resampling (Maris & Oostenveld, 2007). The test operates under the exchangeability null hypothesis, which posits that if no difference exists between two conditions (e.g., healthy and damaged), the group labels are interchangeable. To generate the null distribution, condition labels are randomly shuffled across samples over *M* permutations. For each permutation, the cluster-forming procedure is repeated, and the maximum cluster statistic is recorded. For a given random permutation $m \in 1,2,\dots,M$, the algorithm evaluates all emergent clusters and exracts the maximum absolute cluster statistic, $S_{max}^{(m)}$:

$$S_{max}^{(m)} = \max_k \left|S_k^{(m)}\right| \quad (3)$$

By iterating this process across *M* permutations (where typically $M \geq 1000$), an empirical null distribution $F(S_{max})$ is constructed. This distribution encapsulates the maximum expected spectral deviation under the null hypothesis of exchangeability, implicitly accounting for the spectral covariance structure of the vibration signature. Because the null distribution is constructed directly from the observed data, CBPT is non-parametric and requires no distributional or normality assumptions, accommodating the non-stationary vibration signatures typical of maritime machinery.

Note that equation (3) evaluates the maximum of the absolute cluster statistic, $\left|S_k^{(m)}\right|$, establishing a two-tailed test. As a result, the proposed method detects bidirectional spectral changes, identifying both emergent energy peaks (e.g., fault-induced excitations) and localized attenuation (e.g., increased structural damping or anti-resonances). If diagnostic analysis requires detecting only increased vibration energy, the formulation can be restricted to a one-tailed test by omitting the absolute value operator in equation (3) and evaluating the signed maximum cluster statistic, $\max\left(S_k^{(m)}\right)$, isolating fault-induced spectral peaks while disregarding attenuation.

2.4. Significance Testing and Spectral Localization

Significance is assessed by comparing the observed cluster statistic against the permutation-derived null distribution. Following established permutation testing methods (Phipson & Smyth, 2010), the non-parametric p-value for any observed empirical cluster with statistic S_{obs} is calculated as:

$$p = \frac{1 + \sum_{m=1}^{M} \mathbf{I}\left(S_{max}^{(m)} \geq S_{obs}\right)}{M + 1} \tag{4}$$

where *M* is the total number of random permutations, S_{obs} is the observed statistic for the cluster in question, $S_{\max,m}^{(\pi)}$ is the maximum cluster statistic from the *m*-th permutation, and $\mathbf{I}(\cdot)$ is the indicator function that evaluates to *1* if the condition is true and *0* otherwise. Adding 1 to both the numerator and denominator incorporates the observed data as a valid permutation. This correction guarantees exact control of the FWER and prevents zero-valued p-values, ensuring numerical stability during logarithmic transformations.

An observed frequency band is designated as a critical diagnostic feature if its corresponding p-value falls below a predefined significance level, α. Designating the complete frequency band of a significant cluster as a critical band resolves the MCP while maintaining high sensitivity. This approach identifies the structural footprint of a fault, such as a resonance shift spanning 50 Hz, which point-wise testing often fragments into isolated, non-significant points.

2.5. Mitigating the Large N Paradox via Stability Selection

While spatio-spectral tensors capture the comprehensive dynamics of a mechanical system, processing high-resolution data introduces a statistical vulnerability known as the *Large N Paradox*. In contemporary PHM systems, continuous vibration monitoring generates vast temporal and spectral arrays. As the sample size *N* in frequentist hypothesis testing approaches infinity, the standard error of the test statistic asymptotically approaches zero. Consequently, negligible differences between baseline and degraded mechanical states yield vanishingly small p-values that reject the null hypothesis H_0 . The test therefore detects statistical significance when differences originate from ambient environmental fluctuations or sensor drift rather than genuine machinery degradation.

While CBPT mitigates false positives by clustering adjacent frequency deviations, the adjacency matrix was configured to treat each time-series observation as independent. However, even when restricting cluster formation to the frequency domain, applying this framework monolithically to massive condition monitoring datasets risks hypersensitivity. This statistical overconfidence dilutes the diagnostic value of selected features by obscuring mechanical fault indicators beneath mathematically significant but practically meaningless spectral artifacts.

To isolate robust prognostic features from transient mechanical and environmental noise, the base CBPT framework must be augmented with stability selection. Rather than evaluating the entire $N \times S \times F$ dataset in a single monolithic operation, the proposed methodology executes the permutation test iteratively across multiple randomly subsampled fractions of the training data. This ensemble approach calculates an empirical selection probability, $\Pi_{s,f}$, for every sensor (s) and frequency bin (*f*) based on the proportion of subsampling iterations in which the bin appears within a statistically significant cluster. Setting a high selection probability threshold (e.g., $\Pi_{s,f} \geq 0.95$) filters out spurious anomalies caused by transient noise or localized thermal fluctuations. Consequently, the method retains only spectral deviations that consistently reflect the underlying degradation kinematics, yielding a reliable, low-variance feature set for anomaly detection and diagnosis.

Beyond ensuring statistical robustness, applying the permutation cluster test via fractional subsampling resolves the computational bottlenecks common in multidimensional signal processing. The standard permutation algorithm repeatedly shuffles condition labels and recalculates test statistics over large arrays, consuming substantial memory bandwidth. When aggregated covariance matrices exceed local processor cache limits, execution speed decreases because of continuous, high-latency data retrieval from main memory. Partitioning diagnostic data into smaller, independent fractional subsets keeps active matrices within the L2 and L3 processor caches, accelerating matrix operations. This cache locality reduces computation time and facilitates parallel execution across multiple processor cores, making non-parametric feature selection scalable for fleet-wide prognostic datasets.

2.6. Proposed Methodology: Hierarchical Spatial-Spectral Feature Selection Framework

Univariate feature selection methods, such as analysis of variance (ANOVA), fail to preserve spatial diversity when applied to multi-sensor vibration data. During severe mechanical degradation, structural resonances produce localized, high-energy broadband peaks in the frequency domain. When greedy ranking algorithms evaluate features by point-wise variance, they over-select adjacent, highly correlated frequency bins along the spectral shoulders of a single dominant sensor. To eliminate this redundancy and maintain spatial representation across the sensor array, we propose the *Hierarchical Spatial-Spectral Feature Selection Framework*. This three-stage architecture integrates topologically aware statistical testing with physically constrained peak extraction to generate a robust, pseudo-multivariate diagnostic representation.

2.6.1. Stage I: Statistically-Bounded Spectral Clustering with Global FWER Correction

The first stage uses CBPT to identify regions of statistically significant variance between healthy and faulty kinematic states. To prevent clustering unrelated dynamics across decoupled components, the spatial adjacency matrix is initialized as an identity matrix ($A_{i,j} = 0$ for $i \neq j$), restricting clustering to the frequency axis of each sensor. To control the global FWER across independent sensor evaluations, the Monte Carlo permutation step simultaneously calculates the maximum cluster mass across the entire sensor array. This validates spectral clusters from individual sensors against a unified, globally corrected statistical threshold.

2.6.2. Stage II: Kinematic Centroid Isolation and Topological Prominence

While Stage I isolates contiguous spectral bands containing diagnostic information, broadband leakage often obscures underlying defect frequencies. To extract discrete, physically meaningful features from these clusters, a prominence-driven peak extraction filter is applied to the statistically validated domains.

Stability selection yields a two-dimensional probability matrix $\Pi_{s,f}$, which quantifies the selection frequency of each sensor-frequency bin. Using these raw clusters directly in a prognostic model increases dimensionality unnecessarily, because mechanical modulation produces broad energy shoulders around fundamental frequencies that obscure the resonant peak.

To isolate mechanical kinematics, a centroid extraction algorithm is applied to the probability matrix. First, a global stability threshold, $\tau = 0.95$, filters transient artifacts. The filtered spectral energy, $E^*_{s,f}$, for a given sensor *s* and frequency bin *f* is defined as:

$$E^*_{s,f} = \begin{Bmatrix} \overline{X}_{s,f}, \text{if } \Pi_{\{s,f\}} \geq \tau \\ -\infty, \text{otherwise} \end{Bmatrix} \quad (5)$$

where $\bar{X}_{s,f}$ represents the mean PSD amplitude (in dB) of the evaluation fault condition.

To prevent selecting adjacent spectral bins on the shoulder of a single resonance peak, spectral distance and topographic prominence constraints are applied. Because standard local maxima searches are sensitive to noise fluctuations in the variance floor, a topographic prominence constraint of ≥ 1.0 dB is enforced.

A spectral cluster bin is retained as a kinematic centroid if and only if it is the highest peak between two adjacent valleys and its amplitude exceeds the higher valley by at least 1.0 dB.% This condition rejects local ripples and spectral shoulders, ensuring that extracted features represent distinct mechanical resonances, such as gear mesh sidebands or bearing defect harmonics, rather than spectral leakage. Consequently, a broad structural resonance spanning multiple frequency bins is condensed into a single coordinate representing the kinematic centroid of the mechanical excitation.

The physical frequency of each retained bin is then mapped to the continuous domain:

$$F_{Hz} = F_{start} + (f_c \times \Delta f) \quad (6)$$

where F_{start} is the baseline frequency after truncating low-frequency regions affected by high-pass filter roll-off, f_c is the discrete frequency bin index, and Δf is the constant spectral bin width.

2.6.3. Stage III: Global Cross-Sensor Aggregation and Ranking

In the final stage, the isolated kinematic centroids from all individual sensors are pooled into a global candidate set. Because Stage II eliminated intra-sensor spectral redundancy, adjacent correlated bins from a single sensor cannot dominate the candidate pool. The pooled centroids are ranked by their empirical stability scores (selection probability across bootstrap iterations) and absolute spectral energy and then truncated to a Top-*K* limit. This hierarchical mechanism distributes feature selection across the machine topology, producing a spatially diverse vector that captures distinct fault dynamics across the sensor array.

2.7. Application of CBPT to PHM

CBPT enables automated fault localization and diagnostics. Shipboard machinery signatures rarely occupy a single spectral line; structural resonances, complex gear-mesh harmonics, and cavitation typically manifest as broadband energy shifts. CBPT statistically validates these distributed spectral features without overfitting or missing low-energy precursors. Furthermore, the permutation framework is robust to operational variability in shipboard environments. Because the empirical null distribution is generated from baseline data under specific operating conditions, the test remains valid under varying sensor noise levels, outperforming fixed-threshold alarm systems.

3. EXPERIMENTAL SETUP AND DATA PROCESSING

3.1. Test System and Faults Studied

To evaluate CBPT under realistic operational conditions, this study analyzes condition monitoring data from a multi-stage reciprocating electromechanical system. Unlike the controlled, single-bearing test rigs common in prognostic literature, this testbed exhibits the kinematic coupling, dynamic responses, and thermodynamic variations characteristic of field-deployed machinery. The system architecture comprises multiple interacting components, including prime movers, multi-stage valves, and interconnected kinematic linkages operating under variable load profiles. Consequently, mechanical transmission paths from internal defect sites to surface-mounted transducers are nonlinear. This structural complexity produces a composite signal in which localized fault signatures convolve with the dynamic response of the machine assembly.

The acquired vibration and sensor time-series exhibit pronounced noise contamination and complex non-stationary dynamics. The frequency spectra contain strong amplitude and frequency modulations, transient impacts from reciprocating components, broadband mechanical flow noise, and cross-axis sensor coupling. Conventional parametric condition indicators, which assume stationarity, Gaussian noise distributions, and isolated harmonics, often fail to achieve statistical significance against this dynamic noise floor. In such high-dimensional regimes, CBPT resolves the multiple comparisons problem, isolating contiguous, mechanistically relevant fault clusters from operational background artifacts.

Within this asset, mechanical degradation manifests primarily through three failure modes targeted by the CBPT framework: working fluid leakage, intake pathway obstructions, and thermal management failures.

The first category comprises internal and external working fluid leakage. Cyclic high-pressure loading subjects valve seats and internal components to mechanical stress, increasing susceptibility to seizing or fracturing, which lubricant carryover and carbonization often exacerbate. This degradation, alongside compromised dynamic seals, diminishes volumetric efficiency and discharge pressure (Cui et al., 2009). Furthermore, boundary friction on reciprocating components, such as piston rings and cylinder liners, causes blow-by, reduces effective compression ratios, and contaminates the discharge stream. Externally, casing and fitting leaks develop from corrosive maritime conditions, structural vibration, and assembly tolerances, further reducing system efficiency.

The second fault category involves intake pathway obstructions caused by fouled filtration media or malfunctioning inlet valves. Particulate saturation in intake filters or valve fatigue from debris restricts flow, increasing thermal and mechanical loads on the prime mover and internal stages. This suction restriction attenuates output capacity and accelerates tribological wear across moving parts due to elevated operating temperatures from the additional compression work required to maintain discharge pressure.

The third failure mode involves thermal management degradation within the inter-stage and post-stage heat exchangers. In maritime environments, raw-water strainers prevent debris ingress; however, accumulated organic and inorganic particulate restricts coolant flow. This restriction inhibits heat transfer across the primary thermal interface, causing localized overheating and accelerating microstructural deterioration in critical components. Reduced heat rejection cascades into secondary fluid management failures: insufficient thermal gradients or impaired condensate traps allow moisture to persist in the working fluid, compromising downstream equipment and machine structural integrity.

3.2. Experimental Conditions and Data Acquisition

Data were collected across five operational conditions representing the primary functional states and fault categories of the system: a baseline state and four induced defect states corresponding to the degradation pathways described:

- Baseline: represents normal system function with the machine fully warmed up to steady-state thermal equilibrium.
- Intake pathway restriction (IPR): simulates intake filtration fouling or inlet obstruction.
- Thermal management degradation (TMD): simulates primary coolant obstruction, such as a raw-water strainer blockage.
- Intermediate pressure boundary fault (IPBF): simulates a localized fluid leak within a middle operational stage, replicating progressive internal seal and valve degradation.
- High pressure boundary fault (HPBF): simulates a localized fluid leak in the system's final operational stage, replicating advanced boundary failure under maximum system pressure.

The hardware used to collect data are discussed in depth in Farinholt et al. (2019) and Banks et al. (2022).

Data were collected by alternating sequentially between nominal baseline states and induced fault conditions. Each fault configuration was sustained for 10 minutes after reaching steady-state thermal and mechanical equilibrium. To capture environmental drift, such as ambient temperature fluctuations and barometric pressure variance, baseline data were recorded for 20 to 30 minutes between consecutive fault injections. Four surface-mounted accelerometers continuously acquired vibration time-series throughout the testing protocol, capturing transient startup kinematics and steady-state dynamics at a sampling rate exceeding the Nyquist criterion for all structural resonances of interest.

In initial preprocessing, continuous multi-channel vibration time-series were truncated to isolate steady-state operation and partitioned into non-overlapping macro-blocks. To remove direct current (DC) bias and low-frequency baseline drift from thermal expansion or cable triboelectric effects, the raw response arrays were mean-centered globally and processed using a zero-phase, fourth-order Butterworth high-pass filter. The PSD for each sensor was calculated using Welch's method with localized linear detrending: each macro-block was partitioned into overlapping windows, windowed with a Hann function to suppress spectral leakage, and ensemble-averaged. This averaging reduces the variance of the spectral estimate against the stochastic noise floor characteristic of reciprocating machinery.

Although reciprocating machinery exhibits cyclostationary behavior governed by crank-angle kinematics, the targeted degradation mechanisms, specifically fluid leakage and particulate fouling, manifest predominantly as time-averaged broadband energy shifts and stationary structural resonances rather than angle-dependent transient impacts. To satisfy the wide-sense stationarity assumption, preprocessing excluded transient startup and shutdown periods to isolate steady-state operation. Welch's PSD estimates were then computed using 1-second windows with linear detrending, over which operational speed variations are negligible. This approach captures the macro-degradation signatures while avoiding the computational overhead of cyclostationary demodulation and angle-domain resampling.

To refine the feature space for CBPT, a high-frequency mask discarded out-of-band transducer resonances to isolate the relevant bandwidth. The high-resolution spectral density vectors were then integrated across discrete narrowband intervals using the trapezoidal rule to compute localized band power. Converting raw spectral density into band-averaged power matrices suppressed broadband flow noise and spurious amplitude modulations, producing a low-variance feature matrix for non-parametric statistical evaluation.

3.3. Spatio-Spectral Tensor Formulation and Chronological Stratification

Following sliding-window integration, the combined feature matrices and associated metadata were formatted and stratified temporally.

A secondary DSP step excised the lowest integrated spectral bin. Because this frequency band falls within the stop-band of the high-pass filter, its signal energy is heavily attenuated. In the two-sample Welch's t-test, which provides the test statistic for the permutation framework, this near-zero variance floor collapses the denominator. Uncorrected, this artifact inflates the t-statistic toward infinity, causing the permutation algorithm to detect false-positive clusters in the noise floor. Excising only this zero-index bin eliminates the filter artifact while retaining adjacent bins, preserving kinematic information at the fundamental running speed and primary excitation frequencies of the prime mover.

The filtered two-dimensional feature matrices were reshaped into three-dimensional spatio-spectral tensors with dimensions $N \times S \times F$, representing N temporal observations, S independent sensors, and F discrete frequency bins.

To map experimental runs to theoretical fault families, chronological metadata were parsed across varying fault severities and repeated trials. Except for IPR, which was evaluated at a single restriction level, the protocol included multiple independent runs across medium and high severity levels for each fault category. Parsing the temporal metadata allowed systematic aggregation of all runs and severity levels for a given degradation modality into unified spatio-spectral tensors representing overarching failure modes, such as aggregating all medium- and high-severity stage-boundary leak trials into the IPBF and HPBF arrays.

3.4. Feature Extraction

To quantify the performance of the proposed CBPT framework, its feature selection efficacy is benchmarked against two standard dimensionality reduction techniques: a univariate statistical filter (feature *selection*) and a global orthogonal projection (feature *extraction*). To ensure a fair baseline comparison, PCA was allocated a principal component budget equal to the total size of the global selected feature pool across all fault states, rather than restricting the dimension to the per-state subset size of $K = 15$.

3.4.1. Univariate Analysis of Variance F-Test

The analysis of variance (ANOVA) F-test represents the conventional point-wise approach to spectral feature selection. In this baseline, the multi-channel PSD tensor is flattened, and an independent univariate F-statistic is computed for every discrete frequency bin across all sensors. The test evaluates the ratio of between-group variance (healthy versus faulty states) to within-group variance.

The algorithm ranks bins by their F-scores and extracts the top K features. Although computationally efficient, this baseline illustrates the limitations of point-wise inference. Because the F-test evaluates each frequency bin in isolation, ignoring spatio-spectral adjacency and mechanical coupling, it is susceptible to the MCP. This absence of structural constraints often selects redundant, highly correlated bins clustered around a single dominant harmonic, ignoring secondary fault kinematics. Furthermore, point-wise isolation leaves the F-statistic vulnerable to broadband noise, as isolated noise spikes can artificially inflate localized variance ratios.

3.4.2. Principal Component Analysis

Principal component analysis (PCA) provides the baseline for unsupervised dimensionality reduction. Unlike CBPT and ANOVA, which select discrete, physically interpretable spectral bins, PCA projects the high-dimensional data onto a lower-dimensional orthogonal subspace.

Before decomposition, the flattened spectral data are standardized to zero mean and unit variance to prevent high-energy broadband noise or dominant running speeds from skewing the covariance matrix. To maintain dimensional parity with the feature pools generated by CBPT and ANOVA, the algorithm extracts $K \times M$ principal components (the top K features across M fault states) that maximize global variance across the dataset. Because PCA is unsupervised, it cannot distinguish between deterministic fault variance and stochastic background noise, limiting its efficacy in low-SNR environments. Although PCA eliminates multicollinearity and compresses data dimensionality, the resulting components are linear combinations across the entire input space. This abstraction obscures the underlying mechanical kinematics and severs the physical traceability between diagnostic models and component-level degradation mechanisms, presenting a major limitation for interpretable PHM.

3.5. Robustness Evaluation & Classifier Validation

To evaluate model robustness against harsh shipboard acoustic noise, we injected additive white Gaussian noise (AWGN) into the test datasets. While baseline data captures the mechanical dynamics of the reciprocating testbed, operational maritime environments introduce severe interference that can mask kinematic fault signatures. We adjusted the AWGN power to yield SNRs from 20 dB to -15 dB. Random Forest classifiers trained exclusively on clean baseline data were evaluated on these corrupted signals to assess cross-domain generalization from controlled laboratory settings to noisy field deployments.

To prevent temporal data leakage and evaluate model generalization, the dataset was partitioned chronologically rather than via randomized cross-validation. For each experimental run, the first 70% of the steady-state time series was allocated to the training set, and the remaining 30% was reserved for testing. This chronological partitioning ensures that models forecast future mechanical states strictly from past observations.

A Random Forest classifier (300 estimators, balanced class weights) evaluated the feature sets. Data standardization (zero mean, unit variance scaling) and feature transformations (PCA) were fit exclusively on the clean training partition. The derived parameters were then applied to both the clean and AWGN-injected test sets. This procedure isolates the test data from feature extraction, ensuring that performance metrics reflect out-of-sample generalization in uncharacterized acoustic environments.

4. RESULTS AND DISCUSSION

To validate the CBPT framework, the feature selection efficacy and diagnostic performance of its derived feature set are evaluated using multivariate vibration data from the experimental testbed and benchmarked against PCA and ANOVA baselines. This evaluation assesses the mechanical interpretability, statistical independence, and environmental robustness of the resulting feature spaces in accordance with ISO 13374 (International Organization for Standardization, 2003) guidelines.

4.1. Feature Distribution and Statistical Independence

Figure 1 illustrates the feature selection topography for TMD. The CBPT probability matrix (Π_k) identifies distributed structural resonances across all four sensors. In contrast, the ANOVA F-score matrix isolates highly localized, adjacent frequency bins on the shoulders of single dominant peaks. Figure 2 confirms this spatial diversity for all faults. CBPT selections distribute across the entire machine topology, whereas ANOVA tends to cluster redundantly around isolated frequency bands.

The feature correlation matrices (Figure 3) quantify this redundancy. Features selected via ANOVA exhibit substantial collinearity (mean absolute Pearson r = 0.82, compared with r = 0.40 for CBPT), demonstrating the susceptibility of point-wise selection methods to the MCP. By integrating multivariate adjacency with topographic prominence filtering, the hierarchical CBPT framework yields an orthogonal feature set that maximizes diagnostic information per dimension while minimizing inter-feature correlation.

4.2. Diagnostic Performance and Environmental Robustness

Under ideal steady-state conditions, all three extraction methods adequately separate the fault classes. Figure 4 shows that on unmodified clean data, CBPT, PCA, and ANOVA achieve Macro F1-scores of 0.955, 0.950, and 0.935, respectively. An extraction limit of up to 15 features per fault class provides sufficient dimensionality to diagnose the theoretical failure modes under optimal SNR, even after consolidating redundant features across classes.

The primary advantage of Stability Selection emerges as signal degradation increases. Figure 4 tracks classifier robustness to progressive AWGN injection. PCA performance degrades sharply below 20 dB because the unsupervised algorithm extracts broadband noise as principal components. ANOVA performance collapses below 15 dB as localized noise spikes artificially inflate univariate variance ratios.

CBPT maintains a Macro F1-score above 0.90 down to 5 dB. At the -2 dB threshold, broadband AWGN power exceeds the underlying signal power, masking discrete mechanical resonances across the entire frequency spectrum. PCA and ANOVA fail entirely under these conditions, dropping to F1-scores of 0.148 and 0.326. CBPT preserves a 0.700 F1-score. The empirical null distribution in the permutation framework prevents this elevated broadband noise floor from obscuring localized structural defect signatures.

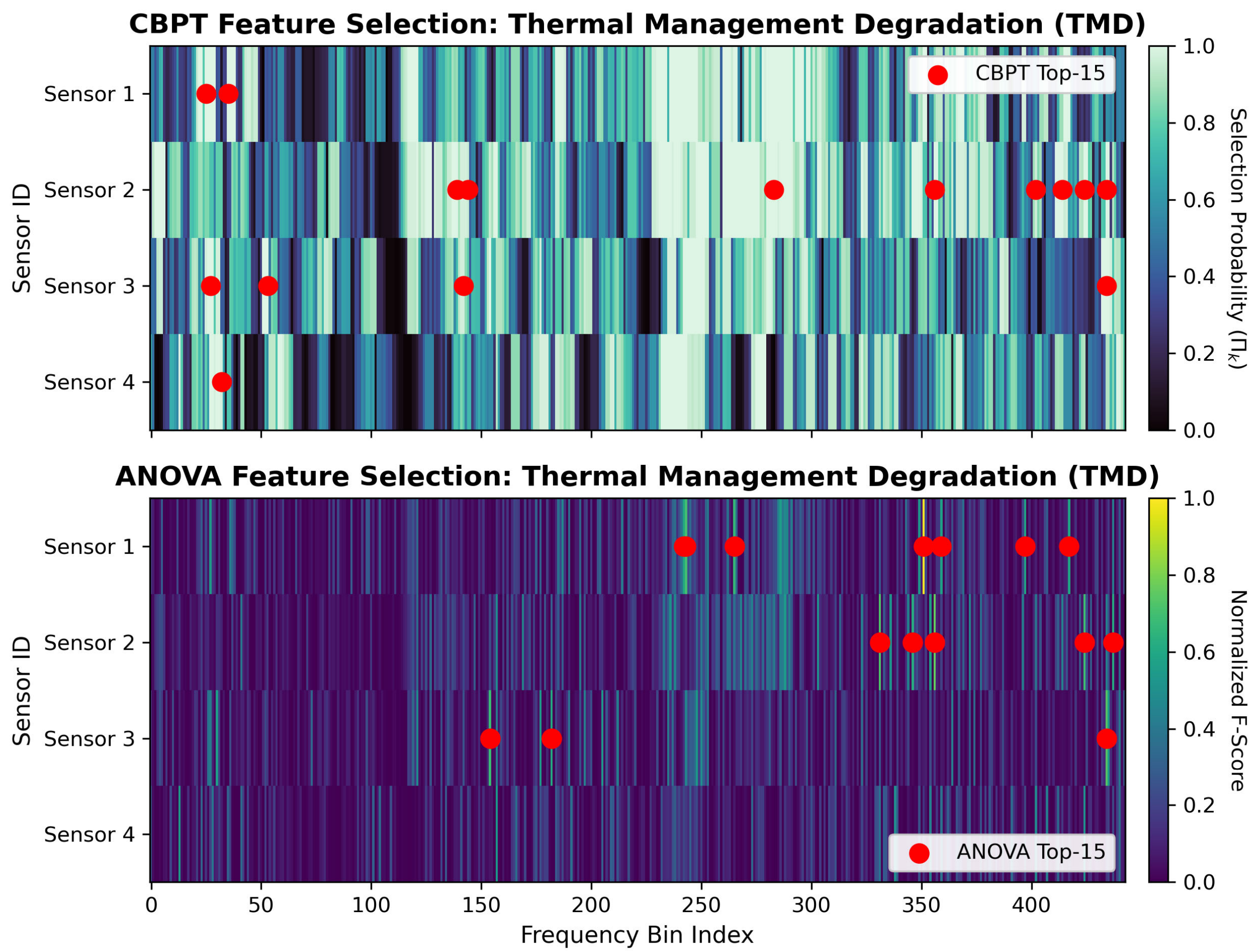

Figure 1. Feature selection heatmaps for Thermal Management Degradation (TMD). The top panel displays the CBPT selection probability across sensors and frequency bins, highlighting distributed resonance peaks. The bottom panel shows the normalized ANOVA F-scores, which cluster heavily around isolated frequency bins. Red circles indicate the top 15 features selected by each method.

5. CONCLUSION

Standard feature extraction methods degrade rapidly in typical naval acoustic environments. Because unsupervised techniques such as PCA maximize global variance, a rising noise floor causes PCA to prioritize stochastic noise over deterministic fault signatures. Similarly, point-wise methods such as ANOVA isolate localized variance ratios but ignore mechanical coupling, selecting redundant, collinear frequency bins along the shoulders of dominant spectral peaks. Consequently, these baseline approaches fail to capture the secondary kinematics required for reliable fault classification under degraded conditions.

The CBPT framework resolves these statistical limitations. By evaluating multivariate spectral adjacency, CBPT identifies contiguous energy shifts corresponding to structural resonances. Constructing an empirical null distribution from baseline data enables the algorithm to distinguish mechanical degradation from ambient noise. Furthermore, integrating fractional subsampling via stability selection mitigates the large-sample paradox, preventing statistical hypersensitivity and ensuring that selected features reflect stable, independent fault kinematics. Sustained classifier performance at negative SNR validates this non-parametric approach for uncharacterized acoustic regimes.

Implementing this hierarchical framework supports embedded health management architectures. Reducing high-dimensional vibration spectra to a defined set of independent kinematic centroids minimizes the computational load on onboard processors while reducing the transmission bandwidth required for shoreside analysis. Future work should evaluate the CBPT framework under transient operating speeds; integrating computed order tracking before permutation testing could extend this methodology to variable-speed propulsion systems.

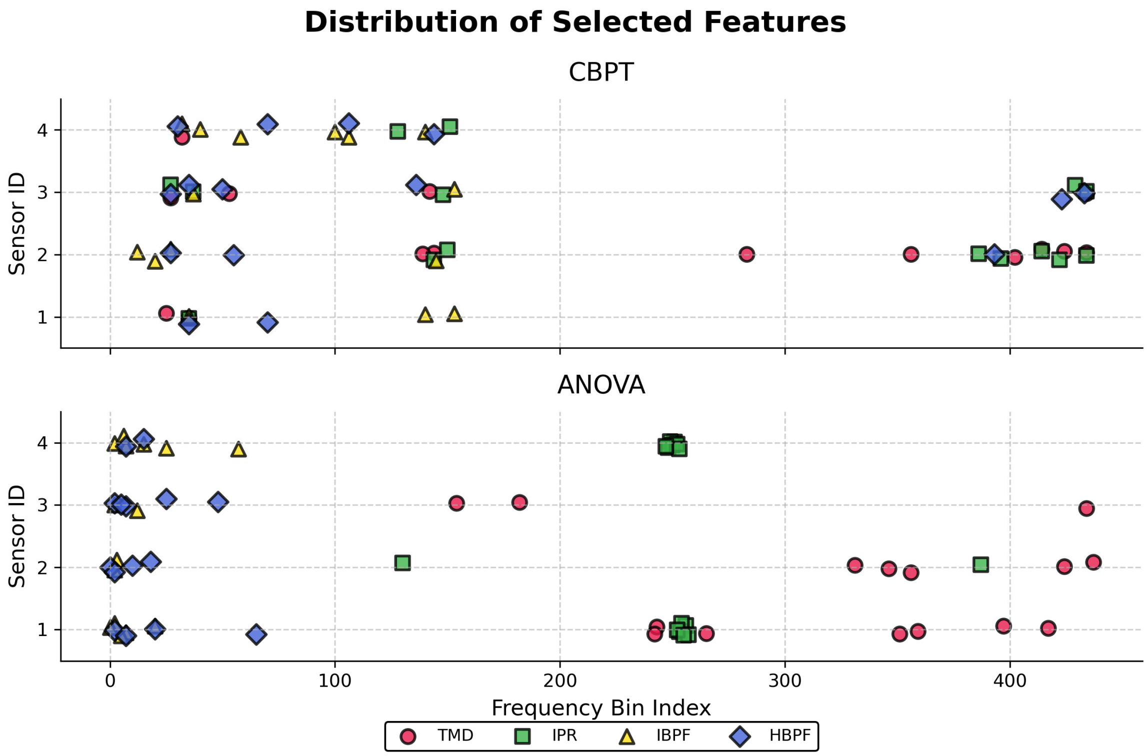

Figure 2. The plot shows the coordinate mapping of the top 15 features selected for all four fault families. The CBPT method selects spatially diverse features across the machine topology. The ANOVA method selects much more correlated frequency bins from isolated sensors.

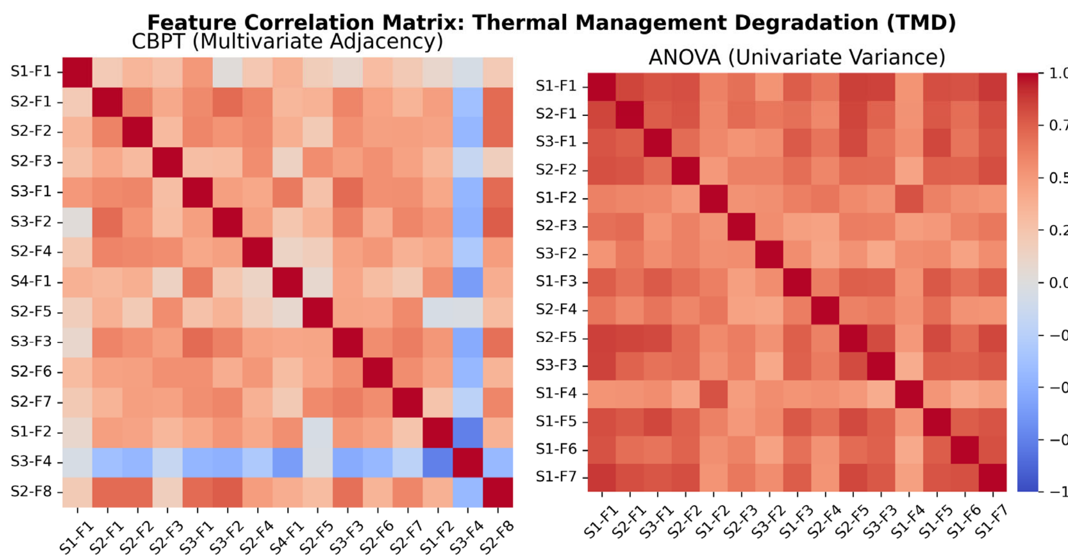

Figure 3. Pearson correlation matrices for TMD by selection approach. The matrices quantify cross-correlation among the top 15 features extracted by CBPT and ANOVA. CBPT features exhibit substantially less cross-correlation, indicating greater statistical independence; mean above diagonal absolute correlation is 0.41 for CBPT, and 0.65 for ANOVA. Both the horizontal and vertical axes denote the selected features grouped by sensor (S_1 to S_4) and indexed by selection order (F_1 to F_n), where S indicates the sensor index (Sensor 1 to Sensor 4) and F indicates feature priority within each sensor.

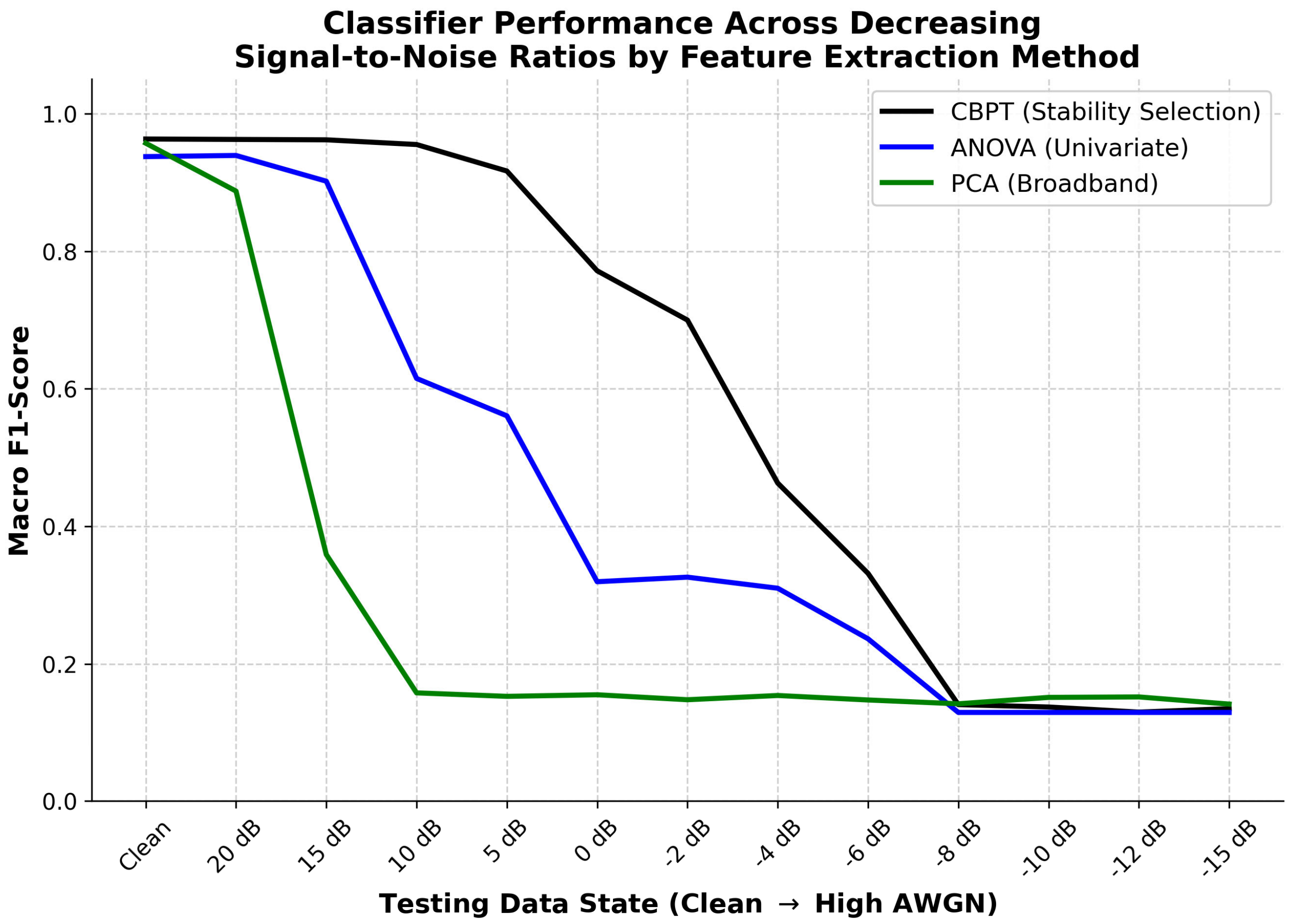

Figure 4. Macro F1-Score degradation across decreasing signal-to-noise ratios. The chart plots classifier robustness to progressive additive white Gaussian noise injection. The evaluation tests models trained on unmodified baseline data against degrade signals. CBPT maintains a Macro F1-score above 0.90 down to 5 dB, outperforming both the univariate ANOVA and broadband PCA baselines under severe acoustic interference.

Although the current methodology evaluates steady-state vibration data, future research will extend the CBPT framework to variable-speed machinery. Integrating computed order tracking or angle-domain resampling prior to permutation testing can synchronize spectral bins with dynamic rotational kinematics.

While this study focuses on discrete fault classification, subsequent investigations will leverage these kinematic features for prognostic modeling. Tracking the amplitude drift of CBPT-selected centroids across extended operational cycles generates monotonic health indicators, providing the state variables needed by particle filters or recurrent neural networks to estimate remaining useful life.

Although environmental robustness was validated against synthetic Gaussian noise, practical field deployment requires benchmarking the algorithm against hull-borne acoustic interference recorded from active vessels. Future work will deploy the feature extraction pipeline on low-power edge hardware to quantify computational latency, cache utilization, and memory bandwidth during continuous real-time monitoring.

ACKNOWLEDGEMENT

The authors are grateful for the Support from PEO SUBS.

REFERENCES

Banks, J., Belella, M., Nixon, S., Polter, S., & Farinholt, K. (2022). *Development of machine learning-based predictive techniques for navy submarine CBM*. Paper presented at the AMTS Conference, 2022.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, *57*(1), 289-300.

Cui, H., Zhang, L., Kang, R., & Lan, X. (2009). Research on fault diagnosis for reciprocating compressor valve using information entropy and SVM method. *Journal of loss prevention in the process industries*, *22*(6), 864-867.

Diaconis, P., & Efron, B. (1983). Computer-intensive methods in statistics. *Scientific American*, 248(5), 116–131.

(DoD) U.S. Department of Defense. (2018). *Condition based maintenance plus (CBM+) for materiel maintenance* (DoD Instruction 4151.22). Washington Headquarters Services.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1), 1–26.

Falzone, S., & Kolodziej, J. R. (2017). Condition Monitoring of a Reciprocating Compressor Using Wavelet Transformation and Support Vector Machines. *Proceedings of the Annual Conference of the PHM Society*.

Fang, Z., Wang, W., Cao, Y., Li, Q., Lin, Y., Li, T., ... & Wu, S. (2024). Reciprocating compressors intelligent fault diagnosis under multiple operating conditions based on adaptive variable scale morphological filter. *Measurement*, *224*, 113778.

Farinholt, K., Polter, S., Hipwell, N., Chaudhry, A., Meekins, R., Kim, M., Thompson, E., & Castillo, K. (2019). Embedded sensing system for condition monitoring of hydraulic actuators. In *Intelligent Ships Symposium 2019 (ISS 2019)*.

Fisher, R. A. (1935). *The design of experiments*. Oliver and Boyd.

Gramfort, A., Luessi, M., Larson, E., Engemann, D.A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L. and Hämäläinen, M., (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroinformatics*, *7*, p.267.

Hou, L., Yi, H., Jin, Y., Gui, M., Sui, L., Zhang, J., & Chen, Y. (2023). Inter-shaft bearing fault diagnosis based on aero-engine system: A benchmarking dataset study. *Journal of Dynamics, Monitoring and Diagnostics*, 228-242.

International Organization for Standardization. (2003). *Condition monitoring and diagnostics of machines — Data processing, communication and presentation — Part 1: General guidelines* (ISO 13374-1:2003).

Jardine, A. K., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, 20(7), 1483-1510.

Lee, J., Barro, A., & Kim, T. (2017). Combustion and Resilient Mounting Condition Diagnostics thru Structural Vibration Monitoring of a Diesel Generator Set. *Asia Pacific Conference of the Prognostics and Health Management Society*.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *Journal of neuroscience methods*, 164(1), 177-190.

Phipson, B., & Smyth, G. K. (2010). Permutation P-values should never be zero: calculating exact P-values when permutations are randomly drawn. *Statistical applications in genetics and molecular biology*, *9*, Article39. https://doi.org/10.2202/1544-6115.1585

Soualhi, A., et al. (2025). Vibration-based Data-driven Fault Diagnosis of Rotating Machines Operating Under Varying Working Conditions: A Review. *International Journal of Prognostics and Health Management, 16(2).*

Trout, J. N., & Kolodziej, J. R. (2016). Reciprocating compressor valve condition monitoring using image-based pattern recognition. *Proceedings of the Annual Conference of the PHM Society*.

Wang, B., Lei, Y., Li, N., & Li, N. (2018). A hybrid prognostics approach for estimating remaining useful life of rolling element bearings. *IEEE Transactions on reliability*, *69*(1), 401-412.

BIOGRAPHIES

Colin Dingley, CISSP, M.S. Cyber Security from Penn State (2024), has spent his entire nine-year career at NSWCPD, becoming an expert in HM&E systems on various naval vessels. He began as a Machinery Control Systems In-Service Engineering Agent, developing and installing the control software sailors use to operate and monitor HM&E equipment. He then migrated to the Unmanned Surface Vehicle program, where he used his HM&E expertise to evaluate how AI/ML based solutions control and manipulate autonomous ships. Now he contributes to this research by leveraging his experience in Machine Learning and HM&E systems to design test environments that allow for proper detection of fault conditions.

Neil Eklund, Ph.D., FPHMS, is a leading expert in Asset Health Management and a seasoned technologist with over 25 years of experience in data science, industrial artificial intelligence, and machine learning. He has led high-impact projects across industries including aerospace, energy, healthcare, and oil & gas, resulting in 16 patents and over 70 technical publications. Dr. Eklund is one of the co-founders the Prognostics and Health Management Society and serves on its board of directors. His career includes significant roles at General Electric Research, Xerox PARC, and Schlumberger, where he was the Chief Data Scientist. At Schlumberger, he led the development of the first successful deployed Internet of Things (IoT) application in the oil industry, generating over $20 million in its first three months of operation. Dr. Eklund's extensive collaboration with organizations like DARPA, NASA, the DoD, Lockheed Martin, ExxonMobil, Ford Motor Company, and Boeing highlights his impact on both commercial and government sectors. His contributions extend beyond industry, having taught graduate-level courses in machine learning and classes through the PHM Society.

Adam S. Wechsler, M.S. in Data Science from Northwestern University, Chicago, IL, USA (2023). He has been involved in data analytics projects with a focus on anomaly and fault detection for Naval Surface Warfare Center Philadelphia

Division (NSWCPD) since 2017. He has generated anomaly and fault detection models for a variety of mechanical equipment as part of the development of an embedded condition-based monitoring data acquisition system. He has also generated anomaly detection and predictive maintenance models from historical ship system and maintenance data. He has recently pivoted some of his effort toward software development for a ship equipment health-monitoring dashboard. Prior to joining NSWCPD, he had worked as a high school mathematics teacher for 16 years, and his involvement in education continues as a co-lead for the SeaGlide high school robotics competition.

Khai Van is an accomplished Mechanical and Electrical Engineer with extensive experience in shipboard system test, evaluation, and platform integration. He possess deep expertise in electrical and mechanical modeling, simulation platforms, and Hardware in the Loop (HIL) testing. His experience spans the development of predictive maintenance health monitoring systems using machine learning algorithms, aiming to optimize maintenance schedules and enhance the efficiency of shipboard systems.

Sherwood Polter, received MSEE in Acoustics and Power Systems Engineering from University of Idaho, is involved with machinery silencing R&D projects since 1996. Mr. Polter is program manager for embedded condition-based monitoring leading to acquiring data in relevant environments for machine learning models to inform predictive maintenance technology development for auxiliary machinery HM&E systems at NSWCPD. Mr. Polter has co-authored papers in CBM applications for the American Society of Naval Engineers, IEEE PHM and PHM Society. Mr. Polter has one patent for resilient mount measurement used in damping and noise control applications. Mr. Polter has led efforts with the development of acoustic noise control applications with design, testing, evaluation and inspection for auxiliary machinery systems.

About NSWCPD The mission of NSWC Philadelphia Division is to provide research, development, test and evaluation, acquisition support, engineering, systems integration, in-service engineering and fleet support. NSWC, Philadelphia Division is responsible for cyber-security, comprehensive logistics, and life-cycle savings through commonality for surface and undersea vehicle machinery, ship systems, equipment and material, and execute other responsibilities as assigned by Commander, NSWC.