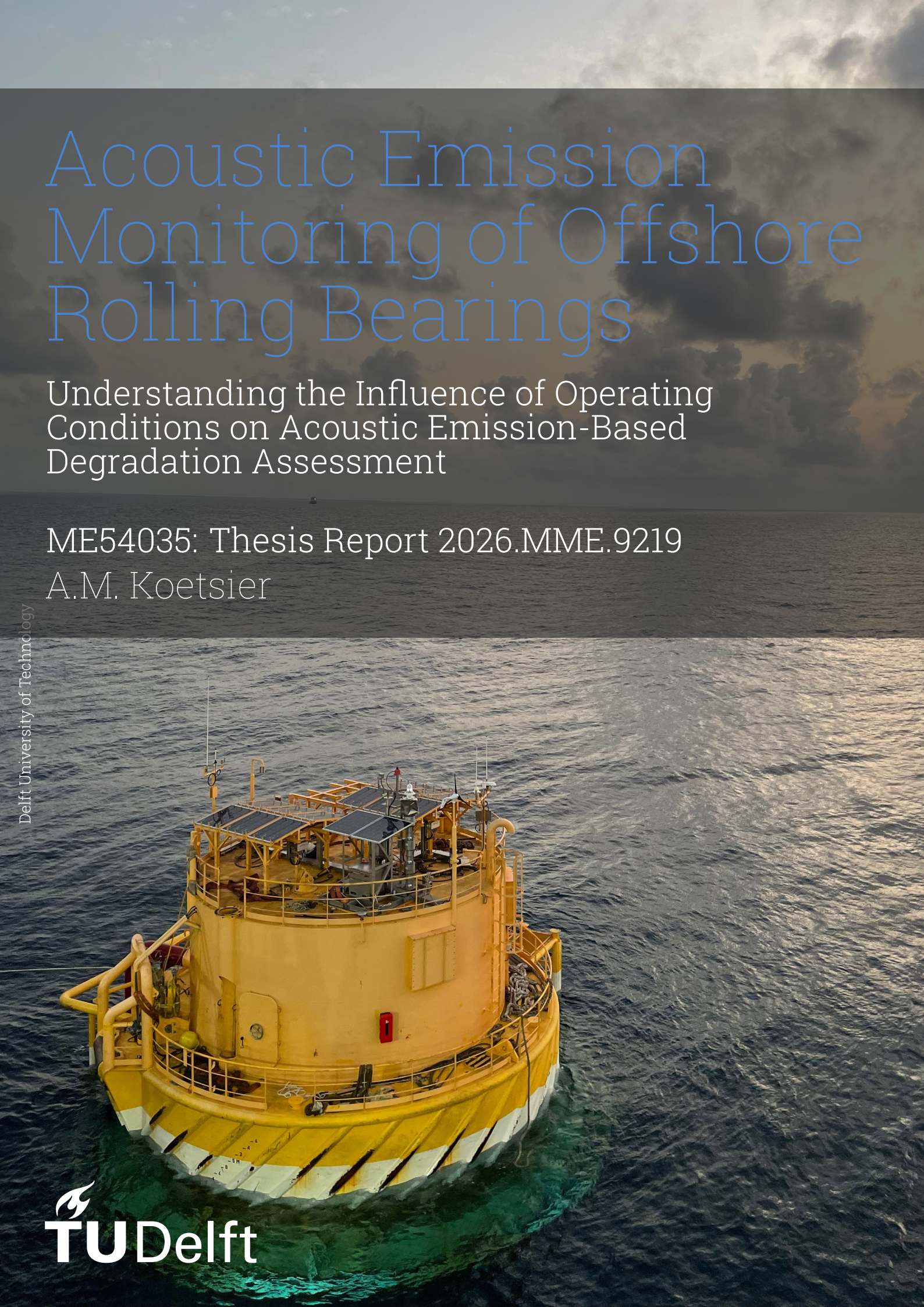
Acoustic Emission Monitoring of Offshore Rolling Bearings

Understanding the Influence of Operating Conditions on Acoustic Emission-Based Degradation Assessment

ME54035: Thesis Report 2026.MME.9219

A.M. Koetsier

Delft University of Technology



TU Delft

Acoustic Emission Monitoring of Offshore Rolling Bearings

Understanding the Influence of Operating Conditions on Acoustic Emission-Based Degradation Assessment

by

A.M. Koetsier

In partial fulfilment of the requirements for the degree of

Master of Science
in Mechanical Engineering

At the Department of Maritime and Transport Technology of the Mechanical Engineering Faculty of the Delft University of Technology.
To be publicly defended on September 28, 2026 at 10:00.

Student Number:	4841948
MSc Track:	Multi-Machine Engineering
Report Number:	2026.MME.9219

Chair:	Dr. J. Jovanova	TU Delft
Supervisor:	Dr. L. Pahlavan	TU Delft
Supervisor:	Dr. B. Scheeren	HiTeAM Systems
Supervisor:	Ir. N. Daane	Bluewater Energy Services

An electronic version of this thesis is available at `http://repository.tudelft.nl/`

Cover:	CALM Buoy by Bluewater Energy Services (Modified)
Style:	TU Delft Report Style, with modifications by Daan Zwaneveld

TU Delft

Samenvatting

Wentellagers in offshore toepassingen draaien zeer langzaam en onder hoge, wisselende belasting. Conventionele trillingsmonitoring is onder die omstandigheden minder effectief. Akoestische emissie (AE) is een alternatief, omdat deze techniek reageert op lokale processen in het lagercontact. AE-kenmerken veranderen echter ook door belasting, snelheid, smering en de meetopstelling. Een verandering in een kenmerk hoeft dus geen schade te betekenen. Dit onderzoek gaat na hoe de invloed van de bedrijfstoestand kan worden beschreven en verminderd. Pas daarna kan de resterende variatie in relatie tot de conditie van het lager worden geïnterpreteerd.

Eén werkwijze is toegepast op vier datasets. Het gaat om een sterk gedegradeerd lineair lager, een intact lineair lager, een lineair lager met toenemende vetverontreiniging en een zwenklager op ware grootte zonder bekende schade. Per frequentieband is eerst gekeken of een kenmerk samenhangt met de gemeten belasting en snelheid. Samenhang alleen was niet genoeg. Een kenmerk werd pas gecorrigeerd als de richting van de relatie stabiel bleef, de relatie zich bij herhaalde metingen liet reproduceren, binnen een goed bemeten bereik viel en voorspellend bleek voor metingen die niet waren gebruikt om het model op te stellen. De correctie zelf bestaat uit een eenvoudig regressiemodel, geschat op een referentietoestand. Daarna zijn de ruwe en gecorrigeerde gegevens met elkaar vergeleken.

De bedrijfstoestand blijkt de variatie in AE-kenmerken sterk te beïnvloeden. Hoe sterk verschilt per kenmerk, frequentieband, sensorpad en lagerconditie. Verontreiniging verandert bovendien de operationele relaties zelf. De correlatie tussen belasting en acquisitie-RMS neemt bijvoorbeeld af van $-0{,}93$ bij schoon vet tot $-0{,}07$ bij de hoogste verontreiniging. Een correctie die voor één smeringstoestand is bepaald, kan daarom niet zonder meer naar een andere toestand worden overgedragen. Binnen de volledige correctieworkflow voldeden uiteindelijk alleen de mediane amplitude en de golfvorm-RMS aan alle eisen. Hun correctie is onderbouwd voor drukken van 0 tot 15 barg en geldt daarbuiten niet.

Ook het intacte lager laat een duidelijke ordening naar bedrijfstoestand zien, zonder dat de conditie verandert. Patronen in ruwe AE-gegevens kunnen daarom niet zonder meer aan schade worden toegeschreven. Na correctie op de intacte referentie neemt de ordening naar snelheid sterk af, terwijl de samenhang met het verontreinigingsstadium duidelijker wordt. Het aandeel van de variatie dat met dit stadium samenhangt stijgt van 0,269 naar 0,468. De scheiding blijft echter beperkt. Een deel van de invloed van de bedrijfstoestand blijft aanwezig in hoofdcomponenten die minder van de totale variatie verklaren. Ook overlappen de stadia nog steeds en kan de invloed van de meetvolgorde niet van het verontreinigingsniveau worden gescheiden.

Het onderzoek laat zien dat de voorspelbare invloed van de bedrijfstoestand meetbaar en deels te corrigeren is. De resultaten ondersteunen daarmee een monitoringsaanpak waarin veranderingen worden beoordeeld ten opzichte van de eigen, operationeel gecorrigeerde referentie van het lager, in plaats van op basis van ruwe AE-kenmerken. Het onderzoek levert geen gevalideerde schadeclassificatie op. Daarvoor zijn metingen nodig aan hetzelfde lager in gezonde en geleidelijk verslechterende toestand, met dezelfde instrumentatie en een gebalanceerde testvolgorde.

AI Statement

For this thesis report for the course ME54035, including the research paper in Appendix A, I have used Generative AI to:

- Obtain inspiration for the overall structure of the report and the research paper.
- Improve the grammar, style, layout and/or spelling of the text.
- Support the development of the Python analysis workflow used in this thesis, including assistance with writing, debugging, restructuring and documenting code for data cleaning, synchronisation, feature extraction, operational characterisation, correction and principal component analysis. All generated suggestions and code fragments were reviewed, tested and adapted by the author, who remains fully responsible for the final implementation and the results obtained with it.

Generative AI was not used to generate, alter or interpret measurement data or results. In all cases, I have critically reviewed and corrected the AI-generated content and remain fully responsible for the content of this report.

Nomenclature

Abbreviations

Abbreviation	Definition
AE	Acoustic Emission
BTF	Bluewater Test Facility
CALM	Catenary Anchor Leg Mooring
EHL	Elastohydrodynamic Lubrication
FFT	Fast Fourier Transform
FLAC	Free Lossless Audio Codec
HF	High-Frequency Band
HiTeAM	High-Tech Acoustic Monitoring
IQR	Interquartile Range
LF	Low-Frequency Band
MF	Mid-Frequency Band
PC	Principal Component
PCA	Principal Component Analysis
RCF	Rolling Contact Fatigue
RMS	Root Mean Square
SCADA	Supervisory Control and Data Acquisition
SHM	Structural Health Monitoring
SNR	Signal-to-Noise Ratio
TB2	TestBlock 2
TB3	TestBlock 3
TRAI	Transient Recorder Index
t-SNE	t-Distributed Stochastic Neighbor Embedding

Symbols

Symbol	Definition	Unit
A	Peak acoustic-emission amplitude	$\mu\mathrm{V}$
$A(m)$	Magnitude of the discrete spectrum in bin m	sample units
A_{dB}	Acoustic-emission amplitude relative to $1\,\mu\mathrm{V}$	dB
A_{eff}	Effective hydraulic area of the BTF swivel bearing	mm^2
$A_{\max}$	Acquisition amplitude ceiling	$\mu\mathrm{V}$
$A_{\mathrm{peak,noise}}$	Peak amplitude of the pre-trigger noise window	$\mu\mathrm{V}$
$A_{\mathrm{peak,signal}}$	Peak amplitude of the complete stored waveform	$\mu\mathrm{V}$
$A_{\mathrm{th},b}$	Amplitude-screening threshold for frequency band b	dB
$\mathcal{B}$	Set of intact baseline windows used to fit the operational model	–
C	Ring-down count of a threshold-triggered AE event	–
C_n	Number of load cycles	–
$C_{n,\mathrm{AE}}(m)$	Load-cycle position of retained hit m	–
C_p	Contamination or particle-content state	–
d	Symmetric relative difference between repeated feature values	$\%$
D	Duration of a threshold-triggered AE event	$\mu\mathrm{s}$

Symbol	Definition	Unit
$\hat{D}(\omega)$	Sensor coupling and detection effects in the source–path–sensor model	–
D_{cond}	Bearing degradation or condition state	–
$D_{\max}$	Acquisition duration ceiling	µs
D_{pcd}	Raceway pitch-circle diameter	mm
D_{seal}	Effective hydraulic seal diameter of the BTF swivel bearing	mm
E_{acq}	Energy recorded by the acquisition system	acquisition units
$E_{f_1:f_2}$	Spectral energy within the frequency interval f_1–f_2	sample^2
F	Applied mechanical load	N
$f_j(\cdot)$	Fitted operational reference model for feature j	feature-dependent
F_z	Measured or hydraulic separating force	kN
f_c	Power-weighted spectral centroid	Hz
f_m	Frequency of spectral bin m	Hz
f_{95}	95% spectral roll-off frequency	Hz
g_2	Biased excess kurtosis of a waveform	–
$h_{\min}$	Minimum lubricant film thickness	m
K	Number of repeats of a varied operating axis	–
L_i	Load step of window i	–
M	Width of the moving-mean smoothing window applied to operational signals	samples
M_h	Number of retained hits in a counting operation	–
m_{95}	Smallest spectral bin index reaching 95% of cumulative power	–
N	Number of samples in a stored waveform	–
N_h	Number of AE hits	–
n_g	Number of observations in group g	–
N_{stroke}	Number of machine strokes	–
$\hat{P}(\omega)$	Measured frequency-domain AE response	–
$\hat{P}_N(\omega)$	Noise, scattering and neglected propagation paths	–
$\mathcal{P}$	Set of primary analysis windows used for scaling and decomposition	–
p_g	Hydraulic gauge pressure	barg
$P_{f_1:f_2}$	Fraction of spectral power within the frequency interval f_1–f_2	–
P_i	Speed-profile phase of window i	–
R^2_{group}	Proportion of retained score-space variation associated with between-group centroid differences	–
$R_{q,1}$, $R_{q,2}$	Root-mean-square surface roughnesses of the two contacting bodies	m
$r_{D/C}$	Duration-to-count ratio, the mean interval between threshold crossings	µs
$r_{E/D}$	Waveform-energy-to-duration ratio	sample^2/µs
r_j	Residual of feature j after operational correction	feature-dependent
r_k	Spearman correlation obtained for repeat k	–
$\mathrm{RMS}_{\mathrm{acq}}$	Acquisition RMS, the RMS of the background noise preceding the threshold crossing, excluding the hit signal	µV
r_s	Spearman rank-correlation coefficient	–
$\bar{r}_s$	Mean Spearman correlation across repeated operating sweeps	–
S	Sensor, coupling and transmission-path effects	–

Symbol	Definition	Unit
$\hat{S}(\omega)$	Emitted AE source signal	–
s_r	Sample standard deviation of repeat-specific Spearman correlations	–
$s_{v,\mathcal{T}}$, $s_{p,\mathcal{T}}$	Population standard deviations of speed and pressure in the training fold	$^\circ\,\mathrm{s}^{-1}$, barg
$\mathrm{SNR}_{\mathrm{lin}}$	Linear signal-to-noise ratio	–
$\mathrm{SNR}_{\mathrm{dB}}$	Signal-to-noise ratio expressed in decibels	dB
T	Temperature	$^\circ\mathrm{C}$
$\hat{T}(\omega)$	Transmission across structural interfaces	–
$T_{0,j}^{\mathrm{AE}}$	Timestamp of the first AE hit in test j	s
$T_{i,j}^{\mathrm{AE}}$	Timestamp of AE hit i in test j	s
$T_{j,k}^{\mathrm{log}}$	Logged start time of pass k in test j	s
$\mathcal{T}$	Set of cells in a cross-validation training fold	–
t_{hit}	Timestamp of an individual AE hit	s
$t_{\mathrm{start},w}$, $t_{\mathrm{end},w}$	Start and end times of analysis window w	s
v	Stroke velocity	$\mathrm{m\,s}^{-1}$
v_a, v_b	Feature values at two visits to the same commanded state	feature-dependent
v_i	Commanded rotation-speed magnitude of BTF cell i	$^\circ\,\mathrm{s}^{-1}$
v_i^*, p_i^*	Speed and pressure standardised on the training fold	–
$\overline{v}_{\mathcal{T}}$, $\overline{p}_{\mathcal{T}}$	Mean speed and pressure of the training fold	$^\circ\,\mathrm{s}^{-1}$, barg
v_{op}	Operating speed, angular or translational depending on dataset	dataset-dependent
W	Number of retained spectral bins	–
$\hat{W}(\omega)$	Propagation through the bearing structure	–
x_{AE}	Generic acoustic-emission feature	feature-dependent
$\bar{x}_M(t)$	Moving mean of an operational signal over a window of width M	signal units
x_{RMS}	RMS computed from the complete stored waveform	sample units
x_{crest}	Waveform crest factor	–
x_{energy}	Summed squared waveform amplitude, used to construct derived ratios	sample^2
x_{kurtosis}	Bias-corrected excess kurtosis of the waveform	–
x_j	Value of AE feature j	feature-dependent
$\hat{x}_j$	Predicted value of AE feature j under the measured operating state	feature-dependent
$\tilde{x}_{ij}$	Transformed value of feature j for observation i	–
$Y(m)$	Discrete Fourier transform of a stored waveform at bin m	sample units
$y(n)$	Sample n of a stored waveform	sample units
$\bar{y}$	Mean of a stored waveform	sample units
$\boldsymbol{z}$	Vector of measured operating variables used in the correction model	variable-dependent
$\boldsymbol{z}_i$	Retained principal-component score vector of window i	–
$\bar{\boldsymbol{z}}$	Overall centroid of the retained score space	–
$\bar{\boldsymbol{z}}_g$	Centroid of group g in the retained score space	–
z_{ij}^{raw}	Robust-scaled raw feature value	–
z_{ij}^{res}	Robust-scaled residual feature value	–
β_{0j}	Intercept of the baseline operational model for feature j	feature-dependent

Symbol	Definition	Unit
$\beta_{Lj,l}$	Regression coefficient associated with load level l	feature-dependent
$\beta_{Pj,p}$	Regression coefficient associated with speed-profile phase p	feature-dependent
β_p, β_v	Pressure and speed coefficients of the BTF correction model	feature-dependent
β_{vp}	Speed–pressure interaction coefficient of the BTF correction model	feature-dependent
Δ_{var}	Cross-validated variance reduction	$\%$
ε	Unexplained variation or model error	feature-dependent
λ	Lubricant film-thickness ratio (lambda ratio)	–
$\tau_{j,k}$	Relative logged start time of pass k in test j	min
$\tau_{i,j}$	Relative timestamp of AE hit i in test j	min
ϕ	Sign-consistency fraction of repeat-specific correlations	–
$\Psi(\cdot)$	Unknown combined influence of operating, lubrication, condition and path variables on an AE feature (conceptual)	feature-dependent
ω	Angular frequency	$\mathrm{rad\,s^{-1}}$

Indices and Operators

Symbol	Definition
b	Frequency-band index
g	Group index, denoting a condition or run label
i	Observation, window or hit index
j	Feature index, or test index in chapter 5
k	Repeat index, or pass index in chapter 5
m	Spectral-bin index
n	Waveform-sample index, or number of paired observations in a rank correlation
w	Analysis-window index
$(\cdot)^{(A)}$, $(\cdot)^{(B)}$	Quantity of the additive (A) or interaction (B) correction model
$(\cdot)^{(-p)}$	Prediction from the model fitted without the cells at pressure level p
$\mathbf{1}(\cdot)$	Indicator function, equal to one when its argument is true
$\mathrm{IQR}(\cdot)$	Interquartile range
$R(\cdot)$	Rank of an observation within its sample
$\mathrm{sign}(\cdot)$	Sign operator
$\mathrm{Var}(\cdot)$	Population variance

Contents

Samenvatting i

AI Statement ii

Nomenclature iii

1 Introduction 1

2 Background and Literature 3
2.1 Degradation and Monitoring of Offshore Rolling Bearings . . . 3
2.1.1 Operating Conditions and Degradation Mechanisms . . . 3
2.1.2 Conventional Monitoring and Its Limitations . . . 6
2.2 Principles of Acoustic Emission Monitoring . . . 7
2.2.1 Physical Basis and Sources . . . 7
2.2.2 Signal Propagation, Measurement and Features . . . 8
2.2.3 Application to Rolling Bearing Monitoring . . . 9
2.3 Effectiveness of Acoustic Emission at Low Speeds . . . 10
2.4 Operational Influences on AE Features . . . 10
2.5 Operational Variability and Normalisation in SHM . . . 11
2.6 Literature Synthesis and Research Need . . . 12

3 Methodology 13
3.1 Datasets and Their Analytical Roles . . . 13
3.2 Scope and Assumptions . . . 14
3.3 Constructing Comparable Observations . . . 15
3.4 Feature Extraction . . . 16
3.5 Operational Characterisation . . . 18
3.6 Correction Eligibility . . . 18
3.7 Feature-Specific Correction . . . 19
3.8 Multivariate Residual Analysis . . . 20
3.9 Evaluation . . . 20
3.10 Software Implementation . . . 20

4 Operational Characterisation of AE Features in Linear Bearing Datasets 21
4.1 Chapter Scope and Shared Analysis Strategy . . . 21
4.2 Phase 4: Advanced Degradation State . . . 22
4.2.1 Recorded Data Overview . . . 22
4.2.2 Bearing Condition and Experimental Programme . . . 22
4.2.3 Acquisition and Analysis Constraints . . . 24
4.2.4 Influence of Load . . . 24
4.2.5 Influence of Stroke Speed . . . 27
4.2.6 Influence of Lubrication State . . . 27
4.2.7 Assessment of Supported Operational Relationships . . . 28
4.3 Phase 2 TestBlock 2: Intact Bearing . . . 28
4.3.1 Recorded Data Overview . . . 29
4.3.2 Experimental Programme and Data Quality . . . 29
4.3.3 Withdrawal of the Load Axis . . . 30
4.3.4 Influence of Stroke Speed . . . 30
4.3.5 Ascending and Descending Repeatability . . . 30
4.3.6 Assessment of Supported Operational Relationships . . . 31
4.4 Phase 2 TestBlock 3: Contamination Progression . . . 31

4.4.1 Recorded Data Overview . . . 31
4.4.2 Experimental Programme and Coverage . . . 32
4.4.3 Influence of Load . . . 33
4.4.4 Influence of Stroke Speed . . . 33
4.4.5 Stability Across Contamination States . . . 34
4.4.6 Condition Associated Changes at Comparable Operating States . . . 34
4.4.7 Assessment of Supported Operational Relationships . . . 36
4.5 Cross Dataset Synthesis . . . 36
4.5.1 Distribution of Operational Sensitivity Across the Feature Set . . . 36
4.5.2 Stability, Repeatability and Coverage . . . 37
4.5.3 Implications for Operational Correction . . . 38
4.6 Chapter Summary . . . 38

5 Full-Scale Swivel Bearing: Experiment and Operational Influence 40
5.1 Test Bearing and Setup . . . 40
5.2 Test Programme and Data Preparation . . . 42
5.2.1 Speed-Step Time Alignment . . . 42
5.2.2 Waveform Decoding and Feature Extraction . . . 42
5.3 Recorded Population and Amplitude Screening . . . 43
5.3.1 Amplitude Screening . . . 43
5.4 Operational Sensitivity . . . 44
5.4.1 Pressure . . . 44
5.4.2 Rotation Speed and Exposure . . . 46
5.5 Directional Repeatability . . . 46
5.6 Operational Correction and Eligibility . . . 47
5.6.1 Correction Models . . . 47
5.6.2 Correction Performance and Eligibility . . . 48
5.7 Limitations . . . 49
5.8 Chapter Summary . . . 50

6 Feature Engineering and Exploratory Modelling 51
6.1 Rationale and Scope . . . 51
6.2 Window-Level Feature Construction . . . 52
6.3 Feature Quality and Primary Vector . . . 53
6.4 Exploratory PCA Procedure . . . 54
6.4.1 Preliminary Operational Residualisation . . . 54
6.4.2 Retention and Label Association . . . 55
6.5 Exploratory Results . . . 56
6.5.1 TestBlock-2 Reference . . . 56
6.5.2 TestBlock-3 Raw and Preliminary Residual Structure . . . 57
6.5.3 Principal-Component Loadings . . . 59
6.5.4 Feature-Vector Sensitivity . . . 59
6.5.5 Preliminary Residual Checks . . . 60
6.6 Bounded Interpretation . . . 61
6.7 Chapter Summary . . . 61

7 Discussion 62
7.1 Operational Influence on AE Features . . . 62
7.1.1 Load and Pressure Dependence . . . 62
7.1.2 Speed and Exposure . . . 63
7.1.3 Lubrication and Contamination . . . 63
7.1.4 Feature, Band and Measurement Dependence . . . 63
7.2 From Operational Association to Operational Correction . . . 64
7.2.1 Eligibility and Held Out Correction at BTF . . . 64
7.2.2 Changing Relationships and Limits of Residualisation . . . 64
7.2.3 Why Correction Models Do Not Transfer . . . 65
7.3 Separating Operational and Stage-Associated Variability . . . 65
7.3.1 Raw and Residual Multivariate Structure . . . 65

7.3.2 Interpretive Limits of PCA . . . 67
7.4 Implications for Offshore AE Monitoring . . . 67
7.4.1 Monitoring Strategy . . . 67
7.4.2 A Baseline Referenced Monitoring Procedure . . . 67
7.4.3 Requirements and Failure Modes . . . 68
7.4.4 What the Procedure Does Not Establish . . . 68
7.4.5 Test and Monitoring Design . . . 69
7.5 Methodological Reflection . . . 69
7.5.1 Workflow and Characterisation Choices . . . 69
7.5.2 Statistical Unit and Feature Construction . . . 69
7.6 Threats to Validity . . . 69
7.7 Synthesis and Boundaries of the Thesis Findings . . . 70

8 Conclusions and Recommendations 71
8.1 Answers to the Research Sub-Questions . . . 71
8.2 Answer to the Main Research Question . . . 72
8.3 Contributions . . . 72
8.4 Recommendations for Monitoring and Test Design . . . 73
8.5 Recommendations for Further Research . . . 73

References 75

A Research Paper 79

B Linear Bearing: Supporting Data 80
B.1 Sensor and Channel Configuration . . . 80
B.2 Operational Characterisation Sequence . . . 81
B.3 Recorded Data Overview . . . 81
B.3.1 Phase 4 . . . 82
B.3.2 TestBlock 2 . . . 84
B.3.3 TestBlock 3 . . . 86
B.4 Scalar Load-Sensitivity Matrices . . . 88
B.5 Waveform-Level Load-Sensitivity Matrices . . . 92
B.6 Sensitivity to the Retained Hit Population . . . 92
B.7 TestBlock 2 Withdrawn Load Matrices . . . 93
B.8 TestBlock 3 Recording Map . . . 94
B.9 Speed-Sensitivity Summary . . . 95
B.10 Lubrication-Sensitivity Table . . . 97

C Full-Scale Swivel Bearing: Supporting Data 100
C.1 Complete BTF Test Matrix . . . 101
C.2 Bearing Dimensions . . . 102
C.3 Uncalibrated Pressure Traces . . . 103
C.4 Operator Rotation Logs . . . 104
C.5 Preprocessing Pipeline . . . 105
C.6 Hit Activity Through Each Recording . . . 105
C.7 Amplitude-Screening Sensitivity . . . 105
C.8 Individual Signed Correlations . . . 105
C.9 Directional Repeatability . . . 107

List of Figures

2.1 Schematic representation by Guo et al. [17] of the main components of a rolling bearing, including the inner race, outer race, rolling elements, cage, shaft and housing. 4
2.2 Schematic cross section of the rolling contact zone showing coexisting degradation mechanisms, including lubricant film breakdown, asperity contact, debris indentation, micropitting and fatigue crack initiation. 4
2.3 Degradation progression framework adapted from [4] with approximate monitoring sensitivity windows for each technique. 6
2.4 Schematic illustration by Scheeren et al. [33] of acoustic emission generation and transmission in a rolling bearing. A local AE source at the rolling contact generates stress waves that propagate through the rolling element, raceways and bearing structure before being detected by externally mounted sensors. 8
2.5 Common acoustic emission event features by Shateri et al. [36], including threshold crossing, peak amplitude, rise time, duration, counts and signal energy. 9

3.1 The analysis workflow. Shaded stages form the operational correction. 14

4.1 Sensor node locations (A–F) on the Phase 4 linear bearing rig. 24
4.2 Phase 4 waveform level load sensitivity across the eighteen load sweep tests. 26
4.3 Phase 4 feature response to stroke speed in the 3-3 constant load triplet ($497\,\mathrm{kN}$). . . . 27
4.4 Representative operational relationships illustrating the Chapter 4 stability assessment. 37

5.1 BTF swivel bearing: cross-section (left) and AE sensor-node layout (right). 41
5.2 Hit composition by pressure. 43
5.3 Hit-rate versus pressure. Filled markers indicate the measurements retained for the reduced 0–$15\,\mathrm{barg}$ sweep used in the main analysis. Open markers indicate the excluded 20 and $25\,\mathrm{barg}$ recordings, during which hydraulic leakage occurred. 45

6.1 TestBlock-3 feature construction, from individual AE records to window-level observations. 52
6.2 TestBlock-2 raw PCA, coloured by speed-profile phase (left) and nominal speed (right). Phase 3 is the maximum-speed phase; phases 1 and 5 are the slowest. 57
6.3 TestBlock-3 PCA, raw (top) and residual (bottom), coloured by nominal speed (left) and contamination stage (right). 58

B.1 Operational characterisation sequence. 81
B.2 Phase 4 hits per channel by band. 82
B.3 Phase 4 amplitude distribution by band. 83
B.4 Phase 4 hits per acquisition file. 83
B.5 Phase 4 duration distribution by band. 83
B.6 TestBlock 2 hits per channel by band (logarithmic scale). 84
B.7 TestBlock 2 amplitude distribution by band. 84
B.8 TestBlock 2 hits per acquisition file. 85
B.9 TestBlock 2 duration distribution by band. 85
B.10 TestBlock 3 hits per channel by band. 86
B.11 TestBlock 3 amplitude distribution by band. 86
B.12 TestBlock 3 hits per acquisition file. 87
B.13 TestBlock 3 duration distribution by band. 87
B.14 Waveform-level load sensitivity, 3-2 speed triplet. 92

C.1 Uncalibrated pressure trace, 7 July 2026. 103
C.2 Uncalibrated pressure trace, 8 July pressure sweep. 103

C.3 Cumulative hits by pressure and frequency band. 106

List of Tables

4.1 Experimental structure and analytical role of the linear bearing datasets. 22
4.2 Instantiation of the preprocessing stages of section 3.3 in each linear bearing dataset. . 23
4.3 Phase 4 load sensitivity, pooled across the eighteen load sweep tests. 25
4.4 Phase 4 assessment of load relationships. A parenthesised tick denotes less consistent direction across the eighteen sweeps; † marks features sensitive to the retained hit population. Repeatability is not testable and is omitted. 28
4.5 Phase 2 baseline load and excitation matrix. 29
4.6 Minimum recorded hit amplitude per TestBlock 2 load step on channels 1 and 7. 30
4.7 TestBlock 2 low-band speed sensitivity across retained hit populations. Values are mean Spearman correlations across load repeats, except the final two rows. 31
4.8 TestBlock 2 ascending versus descending repeatability. Values are the median and maximum symmetric relative differences between the two passes across the seven comparable load–speed combinations. 32
4.9 TestBlock 2 assessment of speed relationships. A parenthesised tick denotes marginal repeatability; n.t. denotes not testable. 32
4.10 TestBlock 3 lubrication contamination levels. 33
4.11 TestBlock 3 load sensitivity, pooled across the four contamination runs. Values are mean Spearman correlations with load. Hit-rate is common to both groups and is reported once. 34
4.12 TestBlock 3 speed sensitivity, pooled across the four contamination runs. Values are mean Spearman correlations with stroke speed. hit-rate is common to both groups and is reported once. 35
4.13 TestBlock 3 low band load sensitivity per contamination run. 35
4.14 TestBlock 3 stability of candidate operational relationships across contamination states. Ticks mark relationships that remain stable; crosses mark condition-dependent behaviour. The parenthesised cross marks a relationship that strengthens with contamination rather than collapsing. 36
4.15 Cross dataset synthesis of operational evidence. 39

5.1 BTF recordings used in the analysis. 42
5.2 BTF pressure sensitivity, 0–15 barg. 45
5.3 BTF speed sensitivity, 0–15 barg. 46
5.4 BTF operational-relationship assessment. 49

6.1 Feature families included in the low/mid core vector and the two sensitivity vectors. . . . 54
6.2 TestBlock-2 retained principal components and dominant loadings. The four largest absolute loadings are listed for each component. 56
6.3 TestBlock-3 PCA organisation, raw and residual. 58
6.4 TestBlock-3 retained principal components and dominant loadings in the raw and residual low/mid core spaces. The four largest absolute loadings are listed for each component. 60
6.5 PCA feature-vector sensitivity. 60

B.1 Phase 4 channel matrix. 80
B.2 Phase 2 channel matrix. 81
B.3 Recorded hit population by dataset. 82
B.4 Phase 4 scalar load sensitivity, starvation (groups 1-1 and 1-2). 89
B.5 Phase 4 scalar load sensitivity, full-mixed grease (groups 2-1 and 2-2). 90
B.6 Phase 4 scalar load sensitivity, full-flushed grease (groups 3-1 and 3-2). 91
B.7 Waveform-level load sensitivity at $\mathrm{SNR} \geq 5$. 93
B.8 Waveform-level load sensitivity at $\mathrm{SNR} \geq 10$. 93

B.9 TestBlock 2 scalar load correlations (withdrawn). 94
B.10 TestBlock 2 waveform-level load correlations (withdrawn). 94
B.11 TestBlock 3 file to operating-state mapping. 95
B.12 Phase 4 speed-effect summary across the six test conditions. 96
B.13 Phase 4 lubrication effect relative to lifetime grease. 98

C.1 Complete BTF test programme and dataset selection. 101
C.2 Documented and estimated dimensions of the BTF swivel bearing. 102
C.3 Operator rotation logs, 7 July 2026. 104
C.4 Operator rotation logs, 8 July pressure sweep. 104
C.5 BTF preprocessing pipeline. 105
C.6 Retention under alternative amplitude screens. 105
C.7 Sensitivity of selected relationships to amplitude screening. 106
C.8 Individual pressure-sensitivity estimates, hit features, complete sweep. 108
C.9 Individual pressure-sensitivity estimates, waveform features, complete sweep. 109
C.10 Individual speed-sensitivity estimates, hit features, complete sweep. 110
C.11 Individual speed-sensitivity estimates, waveform features, complete sweep. 111
C.12 Directional repeatability of BTF features (median symmetric relative difference, %). . . . 111

1

Introduction

Rolling bearings are critical components in offshore rotating machinery. Their failure can compromise system integrity and create economic, environmental and safety risks. Bearings in heavy lifting cranes, mooring systems and wind energy applications operate under high loads, intermittent or oscillatory motion and very low rotational speeds. They are also exposed to variable loading, limited inspection access and corrosive marine conditions [1, 2, 3]. These conditions affect contact behaviour, lubrication and wear development. Reliable condition monitoring is therefore important for safe operation.

Bearing degradation does not proceed through a single isolated mechanism. Fatigue, abrasive and adhesive wear, lubrication related surface distress, debris induced damage, pitting and spalling may develop sequentially or simultaneously. The dominant mechanism can also change as surface condition, lubricant state and contact conditions evolve [4, 5, 6]. The measured response can therefore reflect several interacting processes rather than a single defect.

At the low rotational speeds relevant to many offshore bearings, conventional vibration based monitoring becomes more difficult because defect related dynamic responses are weak [7, 8]. Acoustic emission (AE) provides an alternative by measuring transient elastic stress waves generated by processes including crack initiation, plastic deformation, friction and asperity interaction. AE can therefore respond to local contact processes before a measurable vibration response develops [9, 8, 10]. Previous studies have demonstrated its applicability to low-speed bearing monitoring and fault diagnosis [11, 12, 3].

The same sensitivity that makes AE useful also complicates its interpretation. AE features are influenced by bearing condition, but also by speed, load, lubrication state, contact mechanics and the transmission path between the source and sensor [5, 6, 1, 3]. A change in an AE feature therefore cannot be attributed to degradation from its magnitude alone. The expected response under the current operating state must first be considered. Normal environmental and operational variation can mask changes associated with damage, so measured features must be interpreted in relation to the conditions under which they were obtained [13, 14].

Despite extensive research on AE bearing monitoring, operational dependence is often treated as a secondary influence rather than as part of the condition interpretation problem. Much of the existing literature uses seeded defects, controlled operating conditions or isolated operating parameters. This limits the extent to which changes in AE features can be attributed to bearing condition when several sources of variability are present. The central question is therefore not only whether an AE feature responds to degradation. It is also whether its normal dependence on operating state can be characterised well enough for departures from that response to be interpreted.

This thesis addresses that question for low-speed offshore rolling bearings. Speed and load related quantities are recorded for substantial parts of the experimental data, allowing their influence on the AE feature set to be characterised directly. Where a feature shows a sufficiently stable and adequately supported operational relationship, the predictable part of that relationship is reduced. The raw and

corrected multivariate feature spaces are then compared to determine whether operational structure decreases and whether condition associated structure remains.

Several datasets with different geometries, conditions and operating programmes are used. Each dataset is treated according to what it can support. A full scale bearing, with no known damage, provides the main reference for estimating and validating operational corrections. A contaminated laboratory bearing provides repeated condition states on the same rig to examine whether condition associated structure remains after correction. A degraded laboratory bearing is used to characterise operational response in an advanced degradation state.

The aim is not to develop a validated fault classifier. Instead, the thesis evaluates whether operationally informed processing provides a more defensible basis for interpreting multivariate AE measurements. The methodological contribution lies in linking dataset specific operational characterisation, feature specific correction and residual multivariate analysis within one workflow.

Research question

> In low-speed offshore rolling bearings, how can the influence of operating state on acoustic emission features be characterised and reduced so that the remaining multivariate structure can be interpreted in relation to bearing condition?

Research subquestions

1. How do measured operating parameters influence AE features in low-speed rolling bearings, and how is that influence distributed across the feature set?
2. Which relationships between AE features and operating parameters are stable and predictable enough to support operational correction, and over what operating range can such a correction be estimated and validated?
3. After predictable operational effects have been reduced, to what extent does condition associated structure in the multivariate feature space remain distinguishable from residual operating state variability?

The thesis follows these questions in sequence. Chapter 2 reviews bearing degradation, AE monitoring and the influence of operating conditions on condition sensitive features. It also establishes the need for operationally informed condition interpretation. Chapter 3 translates this basis into the methodology used in this thesis, including operational characterisation, correction criteria and multivariate evaluation.

Chapter 4 characterises how load, speed and lubrication affect AE features in the linear bearing datasets and identifies which relationships support correction. Chapter 5 applies the same reasoning to a full scale swivel bearing and develops and validates operational corrections where supported by the measurements. Chapter 6 compares the raw and baseline-residualised feature spaces using principal component analysis. Chapter 7 discusses the implications and limitations of the results. Chapter 8 presents the conclusions and recommendations.

2

Background and Literature

This chapter reviews the background needed to interpret acoustic emission (AE) measurements from low-speed offshore rolling bearings and establishes why the methodology of chapter 3 is required. Section 2.1 discusses offshore bearing degradation and conventional monitoring. Section 2.2 reviews the physical principles and application of AE, and section 2.3 its effectiveness at low speeds. Section 2.4 examines how operating parameters influence AE features. Section 2.5 places this problem within the structural health monitoring treatment of operational variability. Section 2.6 synthesises the literature and states the resulting research need.

2.1. Degradation and Monitoring of Offshore Rolling Bearings

2.1.1. Operating Conditions and Degradation Mechanisms

Large rolling bearings in offshore systems operate under conditions that are unfavourable from both a tribological and a monitoring perspective. Their duty cycle combines high static and dynamic loads, very low rotational speeds and intermittent or oscillatory motion.

In this thesis, low speed denotes the regime in which conventional vibration based diagnostics lose effectiveness. Vibration diagnosis relies on periodic defect impacts exciting structural resonances. It becomes increasingly unreliable below approximately $100\,\mathrm{rpm}$, and purpose built slow-speed systems remain limited below approximately $10\,\mathrm{rpm}$ [8]. Low speed is therefore taken here as operation below approximately $100\,\mathrm{rpm}$, a bound discussed further in section 2.3. The bearings and test campaigns considered in this thesis operate well below it. Large turret bearings in single point mooring systems are usually described by relative ring movement rather than rotational speed because of their large diameter and slow motion [15]. Wind turbine blade bearings operate at approximately $0.5\,\mathrm{rpm}$ to $5\,\mathrm{rpm}$ and offshore slew bearings often below $10\,\mathrm{rpm}$ [16]. Sheave bearings in offshore heavy lifting cranes may reach approximately $40\,\mathrm{rpm}$ [15]. The main components of a rolling bearing are shown in figure 2.1.

Offshore loading is cyclic and variable. Wave action, wind loading, vessel motions and mooring or crane dynamics continuously change load magnitudes and directions, so contact conditions can vary even during nominally constant operation [2, 3, 16]. The marine environment also exposes bearings to high humidity, salt spray and water ingress, creating corrosive conditions around the contact surfaces and lubricant [18]. Installations are often remote, and inspection depends on vessel and crane availability and on suitable weather windows [1, 2].

These conditions influence which degradation mechanisms dominate. Low rotational speeds and high contact pressures impair lubricant film formation and make mixed or boundary lubrication more likely, which promotes direct asperity contact and friction related surface distress [15]. Surface initiated damage, adhesive wear and contamination driven processes may therefore become more prominent relative to classical subsurface fatigue. Rolling contact fatigue remains central, but practical damage also develops through wear, lubrication disturbance, contamination and environmental exposure [4, 19, 20]. Figure 2.2 illustrates how these mechanisms can coexist within a rolling contact zone.

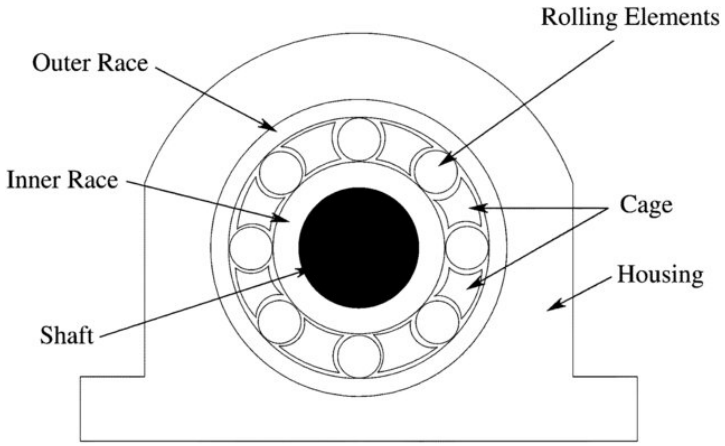


Figure 2.1: Schematic representation by Guo et al. [17] of the main components of a rolling bearing, including the inner race, outer race, rolling elements, cage, shaft and housing.

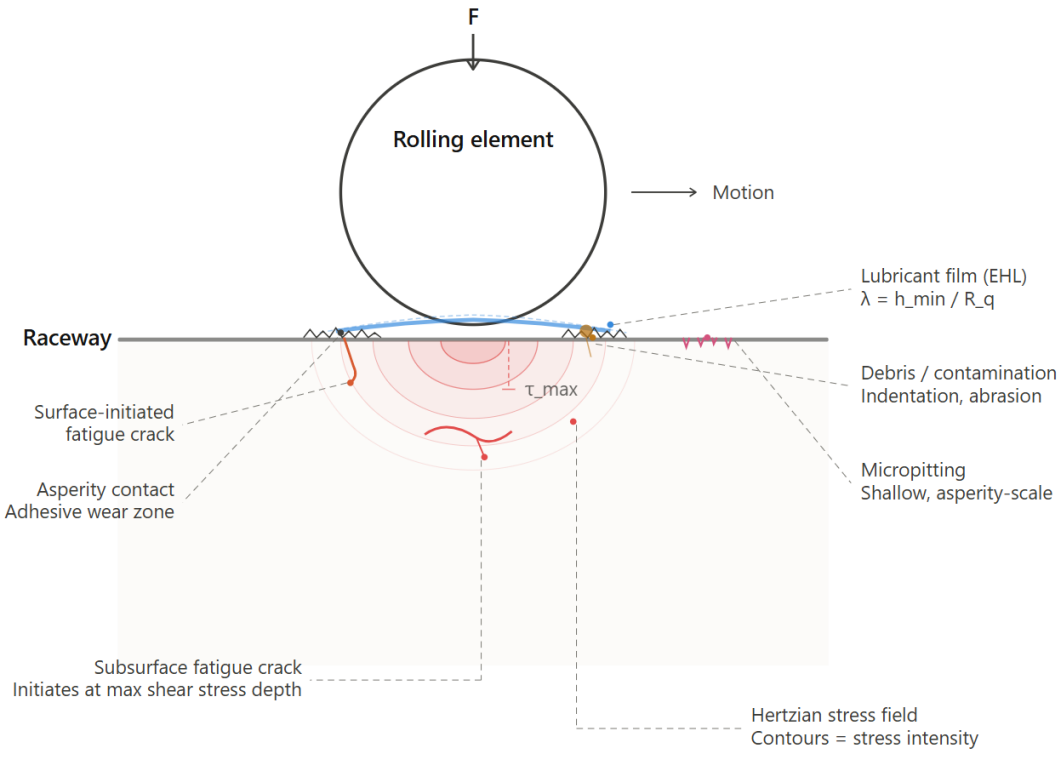


Figure 2.2: Schematic cross section of the rolling contact zone showing coexisting degradation mechanisms, including lubricant film breakdown, asperity contact, debris indentation, micropitting and fatigue crack initiation.

Rolling contact fatigue (RCF) arises from repeated cyclic stresses at the contact interface. Compressive and shear stresses change in magnitude and direction as the rolling element traverses the raceway [21]. Contact pressures may reach several gigapascals, which distinguishes RCF from classical uniaxial fatigue. Damage is commonly categorised into macropitting and micropitting. Macropitting involves larger material loss from propagating fatigue cracks, while micropitting involves shallow distributed surface damage from asperity scale interactions. Both may coexist [22].

Two initiation mechanisms are recognised. Subsurface initiated fatigue occurs under predominantly pure rolling and adequate lubrication. Maximum shear stresses develop below the surface according to Hertzian contact mechanics, with cracks initiating at microstructural stress concentrators and propagating towards spall formation [21, 22]. Improvements in steel cleanliness have made surface condition increasingly influential [23]. Surface initiated fatigue occurs when sliding, lubrication breakdown or geometric irregularities move the maximum shear stress towards the surface, so that crack initiation can occur at asperities, dents or debris induced indentations [22]. The material response under RCF progresses through shakedown, steady state and instability phases [21, 4]. Damage may remain difficult to observe before instability, while the steady state phase can represent a substantial part of bearing life.

Surface wear is particularly important in heavily loaded, low-speed bearings where lubrication is marginal. Abrasive wear removes material when hard particles or raised features plough through a softer counterface. Adhesive wear occurs when insufficient surface separation permits localised bonding, transfer and tearing at asperity contacts [20, 19]. Under oscillatory motion, fretting wear may also develop under small amplitude relative displacement [19]. These wear modes alter surface topography and can promote fatigue crack initiation.

Lubrication effectiveness is characterised by the lambda ratio, which relates lubricant film thickness to composite surface roughness,

$$\lambda = \frac{h_{\min}}{\sqrt{R_{q,1}^2 + R_{q,2}^2}}, \tag{2.1}$$

where $h_{\min}$ is the minimum film thickness and $R_{q,1}$ and $R_{q,2}$ are the root mean square surface roughnesses of the two contacting bodies [24]. Full film elastohydrodynamic lubrication (EHL), in which the raceway and rolling element surfaces are fully separated by a pressurised lubricant film whose thickness is governed by entrainment velocity, lubricant viscosity and the elastic deformation of the contacting bodies, is generally associated with $\lambda > 1$. Below $\lambda \approx 1$, the load is carried predominantly by asperity contacts and boundary lubrication prevails [24]. Because film thickness depends strongly on speed and viscosity, slow or intermittent motion can shift the contact towards mixed or boundary lubrication even without damage. Rycerz et al. demonstrated this at an initial lambda ratio of approximately 0.26, which promoted early surface crack initiation through high asperity stresses [23]. Offshore, low speed limits hydrodynamic lift while high contact pressure reduces the effective film, which can lead to smearing, scuffing and localised heating [1, 23].

Contamination is another persistent concern. Solid particles entrained in the contact create indentations and fatigue crack initiation sites, and particles larger than the film thickness are particularly damaging [24]. Water ingress degrades lubricant performance, promotes corrosion and may contribute to hydrogen assisted embrittlement. Corrosion pits can then act as stress concentrators during subsequent rolling contact [19, 20].

These mechanisms are coupled. Adhesive wear roughens surfaces, fatigue debris intensifies abrasive wear, and contamination induced dents accelerate crack initiation [4, 19]. El Thalji and Jantunen described degradation as a staged progression through running in, steady state operation, defect initiation, defect propagation and damage growth [4]. Degradation may therefore begin long before macroscopic damage is visible. Rycerz et al. reported that crack initiation can occur early in bearing life while much of the remaining life is governed by slow short crack propagation [23]. Techniques sensitive to early stage processes may therefore extend the available warning period. The large variation in life between nominally identical bearings further supports condition based monitoring over life based strategies [22].

2.1.2. Conventional Monitoring and Its Limitations

Several monitoring approaches respond to different consequences of degradation. Vibration analysis is the most established. At moderate to high speeds, localised defects generate periodic impacts that excite characteristic bearing frequencies, which can be identified through spectral analysis, envelope detection or statistical indicators such as kurtosis [24]. At very low speeds, defect impact energy is weak and intervals between impacts are long, so fault signatures become more difficult to distinguish from the normal response. Signal averaging is also less effective because small speed fluctuations prevent repeated impacts from aligning consistently in time [7]. Early stage damage may therefore produce vibration signatures similar to normal variation [1, 24], and clear vibration indicators may only appear once damage is relatively advanced [22].

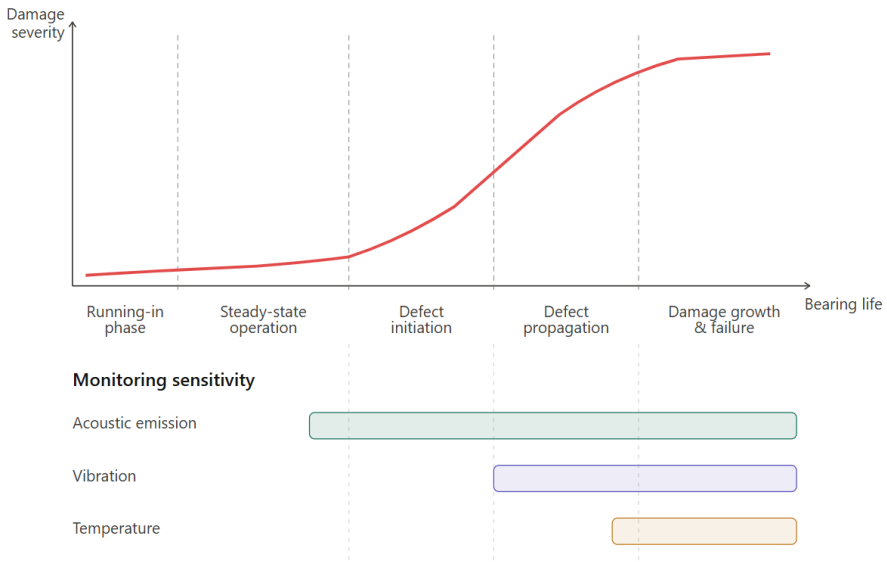


Figure 2.3: Degradation progression framework adapted from [4] with approximate monitoring sensitivity windows for each technique.

Temperature monitoring reflects changes in friction, lubrication and heat generation. It is simple to implement and useful for fault confirmation, but has low sensitivity to incipient degradation, because temperature rise generally occurs after more sensitive indicators have responded [20]. The high thermal mass of large offshore assemblies can delay this response further.

Lubricant and debris analysis provides direct evidence of wear and contamination through particle counting, ferrography and spectrometric analysis [2, 20]. Scheeren et al. showed that abrasive particle contamination increases AE monitoring activity in grease lubricated low-speed bearings, while small or soft particles may be less distinguishable from baseline [18]. Grease lubrication also makes representative sampling difficult. Unlike a circulating oil system, grease does not flow through the bearing, so there is no mixed volume from which a representative sample can be drawn. Wear debris remains close to the location at which it is generated and is distributed unevenly through the grease, so the measured particle content depends strongly on where the sample is taken. Large offshore bearings are also frequently sealed or lifetime greased, so extraction requires physical access and interrupts operation. Debris analysis further depends on material already having been removed, so it may lag the earliest stages of degradation [20, 4].

These methods share several limitations offshore. Each responds primarily to consequences of degradation that appear once damage is already established, so their sensitivity windows open at different points in bearing life, as indicated in figure 2.3. At low speeds, weak defect interactions delay detection [20, 11, 1]. Variable loading, intermittent motion and changing lubrication states can also alter measured signals independently of damage [2, 1]. Large assemblies create complex transmission paths in which local signatures may be attenuated, scattered or filtered before reaching the sensor [1]. Much of the available evidence also relies on seeded defects or accelerated testing, which may not reproduce gradual damage in service [2, 1]. These limitations motivate the use of AE, which is more

closely linked to local contact processes than to the global dynamic response.

2.2. Principles of Acoustic Emission Monitoring

2.2.1. Physical Basis and Sources

Processes such as crack initiation, plastic deformation, frictional sliding and asperity interaction involve rapid release of strain energy within a localised volume. AE monitoring detects the resulting transient elastic stress waves as they propagate through the structure. These waves are measured at the surface using piezoelectric transducers. The distinguishing characteristic of AE measurement is not the transduction principle, which is also used at far lower frequencies, but the observation band, since acquisition is conventionally restricted to frequencies above those that carry most machinery vibration energy [25, 5]. Because the waves originate at or near the material change, AE can contain information about the source event itself. This differs from vibration monitoring, which measures the global dynamic response and generally requires a sufficiently developed defect to alter the system response [5, 12]. AE is also applied more widely in structural health monitoring, for example to crack detection in pipelines [26] and to corrosion and fatigue damage in naval and offshore steel structures [27]. Broader reviews place AE alongside other structural health monitoring sensing techniques [28].

AE signals are commonly classified as burst type or continuous. Burst emissions are short transients associated with discrete energy release such as crack extension or particle impact. Continuous emissions arise from sustained overlapping processes such as frictional sliding or asperity interaction. Both types can occur simultaneously in rolling bearings, and their relative contribution depends on condition, lubrication and operating state.

AE is acquired over a higher frequency band than conventional vibration monitoring. Vibration transducers are nominally sensitive from near DC to approximately $50\,\mathrm{kHz}$, whereas bearing AE instrumentation commonly operates over approximately $20\,\mathrm{kHz}$ to $1000\,\mathrm{kHz}$ [5]. The two ranges overlap at the lower bound, so the distinction is one of analysed quantity rather than a strict frequency separation. Vibration diagnosis interprets the global dynamic response at the defect recurrence frequencies, whereas AE interprets the transient stress waves released at the contact. Defect recurrence frequencies are usually well below $1\,\mathrm{kHz}$ and only a few hertz in low speed bearings, yet each interaction can still generate a short transient with frequency content above $100\,\mathrm{kHz}$ [5, 29]. AE therefore detects the high frequency response produced by individual local events rather than tracking defect recurrence directly.

Field measurements on a mast crane sheave block found a continuous ultrasonic background above $100\,\mathrm{kHz}$ of typically $20\,\mathrm{dB}$ to $25\,\mathrm{dB}$, rising to approximately $35\,\mathrm{dB}$ during operation [30]. Damage related AE occurs as transient activity superimposed on this background, and crack initiation, propagation and spalling are generally associated with more distinct burst type emissions and with increased peak amplitude, counts or energy [12, 31, 11]. Absolute amplitudes cannot be compared directly between studies, because they depend strongly on the sensor, coupling, gain, transmission path and detection threshold. The relevant quantity is therefore the separation between damage related activity and the operating background. The approximately $10\,\mathrm{dB}$ shift observed on going from idle to operation is of the same order as the margin by which discrete emissions typically exceed a detection threshold, which indicates that operating state can reduce this separation rather than acting as a negligible offset.

In rolling bearings, several mechanisms generate AE near the contact. Under mixed or boundary lubrication, roughness features interact directly and generate high frequency stress waves through deformation, shearing and frictional dissipation [29, 32]. Asperity interaction can produce quasi continuous emission to which energy related features such as root mean square (RMS) amplitude respond strongly. Microcrack formation, crack extension and spall development generate more distinct burst type emissions with higher peak amplitudes. Yoshioka, as reported by Elforjani and Mba, used AE to detect subsurface fatigue cracks approximately $50\,\mu\mathrm{m}$ to $200\,\mu\mathrm{m}$ below the raceway [12]. Additional sources include plastic deformation and debris interaction [24].

These mechanisms may be active simultaneously, including under healthy conditions. An increase in AE energy could therefore result from crack propagation, from stronger asperity interaction caused by lubrication changes, or from debris entrainment. No single AE feature can be assumed to represent one mechanism directly.

2.2.2. Signal Propagation, Measurement and Features

Once generated, AE waves propagate through the bearing and the surrounding structure. During propagation they undergo attenuation, dispersion, reflection and mode conversion, so the recorded waveform is not a direct representation of the source. Scheeren et al. described the measured response as the combined effect of the source signal, structural propagation, interface transmission, sensor coupling and background noise [1]. In simplified frequency domain form,

$$\hat{P}(\omega) = \hat{D}(\omega)\,\hat{W}(\omega)\,\hat{T}(\omega)\,\hat{S}(\omega) + \hat{P}_N(\omega), \tag{2.2}$$

where $\hat{P}(\omega)$ is the measured response, $\hat{S}(\omega)$ the emitted source signal, $\hat{W}(\omega)$ propagation through the bearing structure, $\hat{T}(\omega)$ transmission across interfaces, $\hat{D}(\omega)$ sensor coupling and detection effects, and $\hat{P}_N(\omega)$ noise, scattering and neglected paths [1]. The same source can therefore produce different measured responses in different rigs or sensor arrangements. Figure 2.4 illustrates this concept.

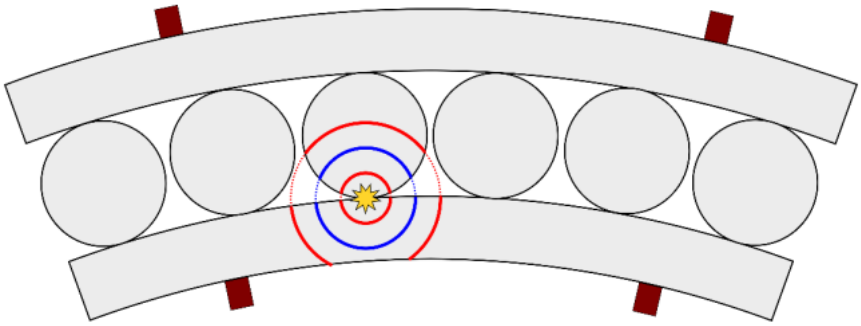


Figure 2.4: Schematic illustration by Scheeren et al. [33] of acoustic emission generation and transmission in a rolling bearing. A local AE source at the rolling contact generates stress waves that propagate through the rolling element, raceways and bearing structure before being detected by externally mounted sensors.

In large assemblies, transmission paths may be long and structurally complex, and high frequency components are particularly susceptible to attenuation. Scheeren and Pahlavan reported that AE signals transmitted through a cylindrical roller bearing remained detectable, but with amplitude reductions of $20\,\mathrm{dB}$ to $30\,\mathrm{dB}$. Average losses were approximately $10\,\mathrm{dB}$ to $15\,\mathrm{dB}$ per interface, with the most favourable transmission below approximately $200\,\mathrm{kHz}$ [34]. Single ring instrumentation may therefore still detect activity on the opposing ring, but with reduced sensitivity. Detectability depends on source location, frequency content and sensor placement [34, 1].

A sufficiently consistent transmission path can also preserve similarities between waveforms generated by similar sources [1, 15]. This provides the basis for waveform similarity analysis, in which resemblance between measured waveforms is used as evidence of a shared source mechanism. The approach reduces reliance on absolute amplitude or spectral content [1, 15], but it assumes sufficient stability of the transmission path over time.

AE sensors are typically piezoelectric transducers mounted on accessible surfaces. Resonant sensors provide high sensitivity within a narrow band, while wideband sensors provide a flatter response at lower peak sensitivity. Reported bearing AE instrumentation spans roughly $40\,\mathrm{kHz}$ to $750\,\mathrm{kHz}$, using either narrowband resonant or broadband transducers [31, 6, 1]. These studies do not define a single standard observation band, so the analysed frequency range depends on the transducer and the measurement configuration. Sensor type, coupling and mounting position influence measured amplitude and spectral content [6], and features obtained in different frequency bands are therefore not directly comparable between studies. This motivates the band resolved treatment used throughout this thesis, with results reported separately by frequency band where the available data provide sufficient coverage.

Signal acquisition also requires high sampling rates, preamplification, bandpass filtering and threshold triggering. The selected threshold affects which events are recorded and therefore influences feature values [1, 35].

Signals can be reduced to descriptive features, derived either from the threshold triggered hit description or from the stored waveform. Figure 2.5 illustrates common descriptors of a threshold triggered AE event. RMS level and signal energy describe overall activity and respond to both continuous and burst type emission, so they cannot directly separate damage related changes from operational changes. Peak amplitude is more responsive to discrete high energy events, while kurtosis is sensitive to impulsive events within a continuous background.

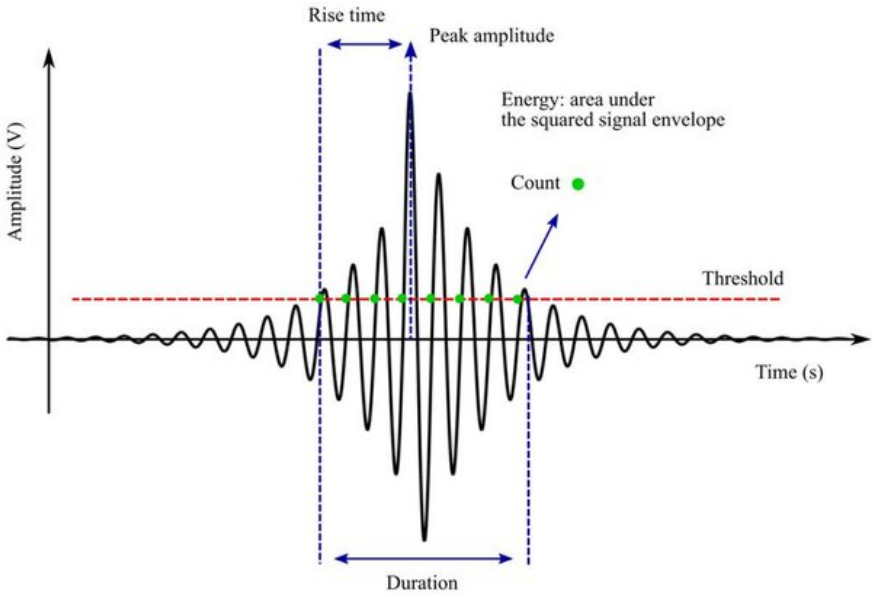


Figure 2.5: Common acoustic emission event features by Shateri et al. [36], including threshold crossing, peak amplitude, rise time, duration, counts and signal energy.

Event based features include hit count, ringdown count and event duration. Tandon and Nakra found that ringdown counts and peak amplitude provided stronger defect detectability than several alternatives [11]. These quantities still depend on the detection threshold and the background noise level [25, 11]. Frequency domain features describe how energy is distributed across frequency bands. Cockerill et al. linked energy between $20\,\mathrm{kHz}$ to $60\,\mathrm{kHz}$ to bearing resonance and energy around $330\,\mathrm{kHz}$ to the sensor resonant frequency [5]. Observed frequency content can therefore reflect attenuation, sensor resonance and transmission effects as well as the source [1].

Time frequency methods can separate transient activity from broader friction related emission [37]. Elforjani and Mba observed increasingly clear periodic evidence of damage using a continuous wavelet transform [31]. Spectral kurtosis has also been used to extract defect related AE from background activity [3].

Overall, features such as RMS and energy aggregate activity from several mechanisms and are sensitive to both damage and operating state. More targeted features such as kurtosis or waveform similarity may offer greater discrimination but require more careful interpretation. No single feature is universally optimal.

2.2.3. Application to Rolling Bearing Monitoring

AE has been applied to rolling bearing monitoring in seeded defect tests [11, 38], in naturally developing degradation experiments [12, 31], in very low-speed studies [8, 39] and in large scale configurations [1, 15, 40]. Early studies showed that ringdown count and peak amplitude can detect defects where vibration based indicators become less effective [7, 11]. Elforjani and Mba later showed that AE energy, counts and burst characteristics evolve from running in and crack initiation towards spall formation in slow-speed thrust bearings [12, 31].

More recent studies have examined larger configurations. Scheeren et al. applied waveform similarity clustering to naturally developing damage in a large scale low-speed bearing segment and related re-

peatable AE source patterns to damage progression [1, 15]. Multivariate AE feature representations reduced with principal component analysis have also been used for low-speed bearing fault diagnosis [41]. These studies show that AE interpretation depends not only on damage, but also on sensor access, transmission paths, lubrication, speed, load and environmental noise.

2.3. Effectiveness of Acoustic Emission at Low Speeds

Monitoring becomes more difficult as rotational speed decreases, because defect related interactions become weaker, less frequent and harder to distinguish from background noise [7, 8, 42]. The bounds introduced in section 2.1.1, approximately $100\,\mathrm{rpm}$ for conventional vibration diagnosis and approximately $10\,\mathrm{rpm}$ for purpose built slow-speed systems [8], mark where this loss of effectiveness becomes limiting.

AE is also affected by speed. Caesarendra et al. distinguished three regimes, with impact dominated AE at higher speeds, contact dominated behaviour at moderate low speeds, and friction and rubbing dominated activity below approximately $10\,\mathrm{rpm}$ [3]. The informative features change between these regimes. Impact dominated behaviour produces more distinct bursts that can be captured through peak amplitude and kurtosis. Friction dominated behaviour produces lower amplitude continuous activity that is better represented by energy based features such as RMS. Mba et al. linked this to sliding velocity and contact dynamics, noting that asperity deformation rates and released energy decrease with speed [43]. AE therefore remains relevant at low speed, but the measured signals can approach the noise floor.

Two studies illustrate this sensitivity. Smith compared high frequency acceleration, jerk, shock pulse and AE on a bearing with a small surface scratch across approximately $10\,\mathrm{rpm}$ to $3750\,\mathrm{rpm}$ [7]. With introduced gear drive vibration, most methods no longer produced identifiable signatures below $300\,\mathrm{rpm}$, whereas AE still detected defect related responses down to $10\,\mathrm{rpm}$. Tandon and Nakra compared seven techniques on bearings with spark eroded defects across $100\,\mathrm{rpm}$ to $1500\,\mathrm{rpm}$. AE ringdown counts maintained detectability to the lowest tested speed, and peak amplitude was the only indicator that reliably detected the smallest ball defects [11].

Both studies used seeded defects in small laboratory bearings. Later work at very low speeds confirmed that detection depends on whether the stress wave energy reaching the sensor exceeds the background noise floor [39, 8]. Source energy, transmission losses and background noise therefore remain important.

Naturally developing damage presents a more demanding case. Elforjani and Mba observed AE activity evolving through stages consistent with running in, crack initiation, propagation and spall formation in low-speed thrust bearings [12, 31]. Scheeren et al. monitored naturally developing damage in a purpose built linear bearing segment representative of an offshore turret bearing [1]. hit-rate increased consistently with inspection but did not isolate damage related activity from the wider AE population. Waveform similarity clustering identified repeatable source patterns that were physically consistent with the observed damage [1, 15].

Across the reviewed studies, AE detection has been demonstrated from $1500\,\mathrm{rpm}$ [11] down to approximately $0.5\,\mathrm{rpm}$ [39]. Performance depends on the ratio between defect related activity and background noise, and on defect severity and scale. Much of the evidence still comes from small laboratory bearings, while studies of natural degradation at large scale remain less common [12, 1, 3].

Most studies also use constant or slowly varying operating conditions. Offshore bearings instead operate under variable loads, intermittent motion and changing lubrication states. Liu et al. partially addressed this for a wind turbine blade bearing under time varying low speed and heavy load [16]. The influence of operating conditions on the AE features themselves nevertheless remains important, because feature changes may reflect operating state rather than damage progression.

2.4. Operational Influences on AE Features

The measured AE response is not determined by damage alone. Rotational speed is the most consistently reported operational influence. It affects the rolling element interaction rate, the sliding velocity and the lubricant film thickness [3, 5, 29]. Couturier and Mba linked AE RMS to predicted specific film thickness under EHL conditions, showing that speed changes the tribological state of the contact [29].

Cockerill et al. also found speed to be the dominant influence on AE RMS in cylindrical roller bearings, while load had a weaker effect under full film lubrication [5]. Although AE features generally increase with speed, the functional relationship has not been established consistently across bearing types [5, 29]. Relationships derived at moderate speeds also do not necessarily extrapolate to very low speeds [7, 8, 39].

Load affects AE through contact mechanics. Increasing load raises Hertzian pressure and can reduce effective film thickness, so its influence is both direct and mediated through lubrication [29]. The measured response is less consistent than for speed and depends on lubrication regime, geometry, defect type and feature [5, 44]. Seeded defect studies often show AE energy increasing with load and speed [38], but responses can be nonlinear and count based features may saturate with defect size [25]. In offshore and wind turbine applications, load also varies together with speed, temperature and structural dynamics. Guo et al. showed that AE indicators for a wind turbine main bearing varied alongside several operating parameters [40].

Lubrication mediates several effects of speed and load. As the lambda ratio decreases towards and below unity, asperity interaction becomes more important and friction driven AE can increase without progressive damage [1, 45, 29]. Cornel et al. reported sensitivity to lubrication condition in wind turbine roller bearings, but also found differences between studies, which suggests that lubrication related behaviour is system specific [45].

Grease contamination creates a similar interpretation problem. Miettinen and Andersson found that AE pulse counts detected contaminant concentrations as low as $0.02\,\%$ by weight, with smaller particles generating higher AE levels [46]. Contamination induced increases can therefore resemble damage related increases. Scheeren's run to failure experiments similarly showed that contamination induced AE could influence a proposed bearing condition index [15]. Temperature is less often investigated independently and acts mainly through lubricant viscosity and film thickness [5, 45, 29].

These parameters do not act independently. Their combined influence on an AE feature can be written conceptually as

$$x_{\mathrm{AE}} = \Psi\left(v_{\mathrm{op}}, F, T, \lambda, C_p, D_{\mathrm{cond}}, S\right) + \varepsilon, \tag{2.3}$$

where $\Psi(\cdot)$ is an unknown, generally non-linear function describing their combined influence, x_{AE} is an AE feature, v_{op} the operating speed, F load, T temperature, λ the lubrication parameter, C_p contamination or particle content, D_{cond} degradation or condition state, S sensor and transmission-path effects, and ε measurement noise. This is not a predictive model. It represents the interpretive problem that the same feature can change because of operating conditions, lubrication, contamination, transmission effects or damage. Field evidence confirms that these variables can vary together in practice [40, 6]. Much of the literature instead varies one parameter while holding the others constant, which isolates individual mechanisms but does not represent how they interact in service.

A raw AE feature therefore reflects both operating state and bearing condition. It can only be interpreted directly when operating conditions are controlled or when the feature is known to respond primarily to damage [12, 11]. Otherwise, the feature must be interpreted relative to the operating state in which it was measured.

2.5. Operational Variability and Normalisation in SHM

Sohn reviewed evidence that environmental and operational variability can mask or be mistaken for damage related changes, which makes normalisation necessary for reliable condition interpretation [13]. Figueiredo and Brownjohn placed this within the statistical pattern recognition framework for structural health monitoring, in which operational evaluation, feature extraction and statistical modelling form successive stages, and normalisation is used to interpret features within their operational and environmental context [14].

An important distinction is whether the relevant operating variables are measured. If they are not, their influence must be inferred from the feature data using latent or unsupervised methods, and dominant modes of normal variation are then used as a proxy for operating state. If the operating variables are measured, their influence can instead be modelled explicitly [47].

Worden et al. argued that observations should be compared with reference behaviour from comparable operating conditions rather than with one reference population that combines all conditions [48]. Pooling different operating states increases the apparent normal variability and can hide genuine departures. Their regression approach estimates the expected reference response from measured environmental variables and can include several conditioning variables. Its validity depends on adequate sampling of the operating space, so the reference should not be assumed to hold outside the range in which it was estimated [48].

Principal component analysis (PCA) is widely used in structural health monitoring for visualisation, feature selection, pattern evaluation and normalisation [14]. Some normalisation approaches model dominant directions of normal variation using PCA and examine the remaining space for condition related changes [13, 14]. This has limitations when operating variables are already measured.

First, PCA identifies directions of maximum variance rather than directions that separate operating state from condition, so a condition related change may occur in a low variance direction that is discarded [14]. Second, the number of components removed must be selected. Removing too few can leave operational effects, while removing too many can remove condition information [47]. Nonlinear operational response creates a further difficulty. Yan et al. showed that a global linear PCA model can produce apparent abnormalities when reference behaviour changes nonlinearly with environmental conditions [49]. Structure in a PCA representation therefore cannot automatically be interpreted as condition related.

2.6. Literature Synthesis and Research Need

Three findings from the reviewed literature frame the remainder of this thesis.

First, AE remains sensitive to degradation at low rotational speed, where vibration based methods become less effective. However, AE features also respond to speed, load, lubrication, contamination and transmission conditions. A change in feature magnitude alone can therefore not be attributed to degradation.

Second, the structural health monitoring literature shows that operational variability must be considered before measured changes are interpreted as condition changes [13, 14]. When operating variables are measured, their influence can be modelled explicitly and features can be interpreted relative to the expected response for the current state [48]. Such a correction is only supported over the operating range in which the reference behaviour was estimated.

Third, PCA can represent multivariate structure in AE features, but it does not establish whether that structure originates from operating state or bearing condition [14, 49]. PCA can support operational normalisation, but this requires assumptions about which variance directions represent operating variability. It is therefore most useful after the role of operating state has been considered.

Within the literature reviewed here, operational influences are more commonly examined individually than within a connected workflow of operational characterisation, feature correction and multivariate interpretation. This workflow is developed in the remainder of the thesis. Chapter 3 defines how comparable observations are constructed, how operational relationships are characterised and screened, how corrections are estimated and validated, and how the corrected feature space is examined.

3

Methodology

Chapter 2 established that AE features respond to operating state as well as to bearing condition, and that this operational variability must be considered before condition is interpreted. This chapter defines the workflow used to put these principles into practice. It specifies how comparable observations are constructed, how operational responses are characterised, how features are screened for correction, how corrections are estimated and validated, and how the resulting feature space is examined using PCA.

The workflow is shown in figure 3.1. Raw records are converted into comparable observations and features, and each feature is then characterised against operating state and screened for correction eligibility. Features that pass every criterion are corrected and are represented by their residuals in the subsequent analysis. Features that fail any criterion are retained unchanged as descriptive features. The resulting feature space is then examined for structure that remains after correction. The complete sequence can only be closed where a dataset supports held-out validation, which is the case for the full-scale swivel bearing (chapter 5). Elsewhere, the correction stage is applied as an exploratory residualisation, as for the contamination progression in chapter 6.

3.1. Datasets and Their Analytical Roles

Four datasets are used, acquired on three test rigs. They differ in bearing type, size, sensor configuration, sampling rate and operating programme, and are therefore not interchangeable. Each is used for the analytical role supported by its condition, instrumentation and experimental design rather than being forced through the same procedure. They are referred to throughout by the names below.

Phase 4 is a double linear bearing at the end of its service life, tested in the HiTeAM Phase 4 campaign. Its within-recording load sweeps make it the clearest source of load dependence, but because the bearing is severely degraded it characterises operational response in an advanced degradation state and is not used as a reference condition (chapter 4).

TB2 and **TB3** are two test blocks of the HiTeAM Phase 2 linear bearing rig, sharing its geometry and acquisition chain. TB2 is an intact bearing in the as-built state. It provides the within-recording speed axis and direct repeatability evidence in chapter 4, and the intact multivariate reference in chapter 6. TB3 comprises an in-block uncontaminated baseline followed by three successive grease-contamination stages. It is used to test whether operational relationships remain stable as lubricant condition changes (chapter 4), and whether stage-associated structure remains after baseline-fitted residualisation (chapter 6). Because TB3 has no independent validation structure, its residualisation is treated as an exploratory multivariate analysis rather than a validated correction.

BTF is a full-scale swivel bearing without known damage, tested at the Bluewater Test Facility. Its measured variability is treated primarily as reference operational variability, and its independent pressure recordings allow held-out validation. It is therefore the only dataset on which the complete correction-eligibility sequence, including predictive adequacy, is demonstrated (chapter 5). Because this experi-

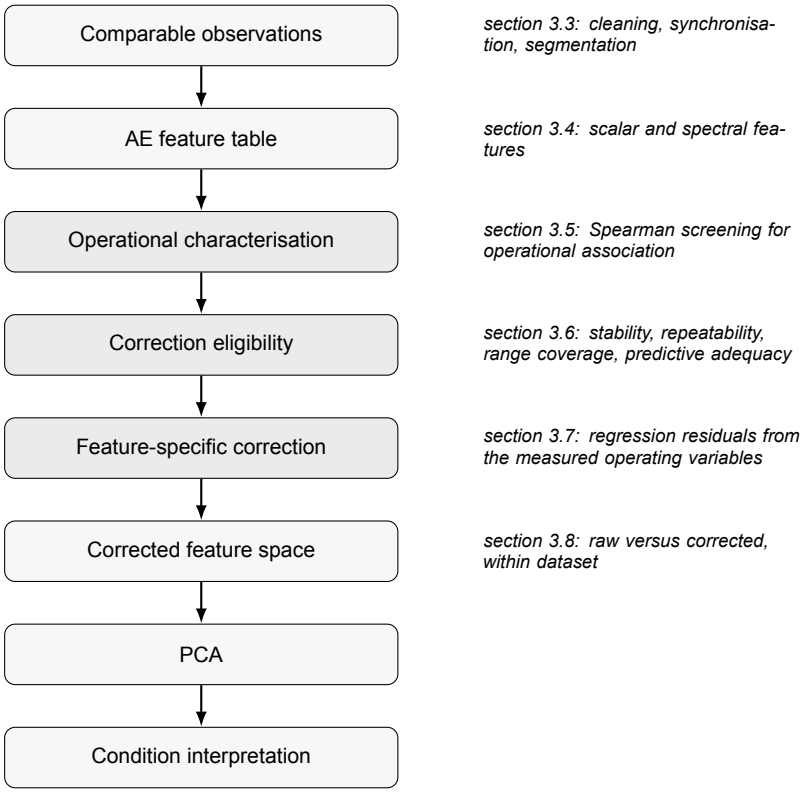


Figure 3.1: The analysis workflow. Shaded stages form the operational correction.

ment was conducted for this thesis, its rig, instrumentation and load generation are described in more detail.

Each dataset is processed independently. Sensor responses and transmission paths differ, so absolute feature values are not assumed to transfer between rigs. The transferable element is the analysis framework, and all comparisons, including those between raw and corrected feature spaces, are made within a single dataset.

3.2. Scope and Assumptions

The recorded operating variables are treated as measured proxies for operating state rather than as a complete description of the bearing contact. Unmeasured influences, including lubrication state, contact distribution and transmission path changes, may remain in the corrected response. The correction targets predictable dependence of the feature average on the measured operating variables. It does not assume that all operational variability is removed.

Consecutive analysis windows are temporally correlated, while condition information is generally available at dataset or experiment level rather than for every analysis window. The analysis is therefore interpreted as exploratory evidence of operational and condition-associated structure rather than as validated diagnostic classification.

The workflow rests on the following assumptions.

- **Heterogeneity as representative variability.** The datasets differ in bearing type, size, sensor configuration, mounting, sampling rate and duration. Analysis is therefore performed within each dataset.
- **AE features as condition-associated signals.** No one-to-one mapping between a feature and a damage state is assumed. Features are influenced by both condition and operating state.
- **Transmission path stability within a dataset.** The source to sensor path is assumed to remain approximately constant during a dataset's measurement period. This assumption weakens if

coupling changes.

- **Operational variables as proxies for operating state.** Recorded variables are assumed to capture the main measured operational variation. Unrecorded factors remain as unexplained residual variation.
- **Lubrication state not directly measured.** No dataset records lambda ratio, film thickness or grease condition in real time. Lubrication effects are therefore inferred indirectly.
- **Approximate stationarity within analysis windows.** Analysis windows are defined to represent one nominal operating state, but stationarity is not guaranteed because measured and unmeasured influences may still vary within a window.
- **Temporal correlation between windows.** Consecutive windows are correlated, so statistical independence is not assumed and formal significance tests are interpreted with caution.
- **Linear operational relationships as a starting point.** Correction begins with linear and low-order interaction models because of their interpretability and the available sample sizes. Nonlinear terms are considered only when residual structure supports them.
- **Absence of per-window fault labels.** Health is generally known at dataset level rather than per window. The workflow therefore does not rely on supervised classification.
- **Sensor frequency response differences.** Absolute spectral content is not assumed to be comparable between datasets. Spectral interpretation is therefore primarily within dataset.

3.3. Constructing Comparable Observations

Because the datasets differ in hardware and format, preprocessing is defined through common analytical stages. Their implementation varies by dataset. Choices such as channel mapping, threshold convention and clock offset are stated in the corresponding dataset chapter. Together, the applicable stages convert the raw AE and operational records into a synchronised feature table.

Validation and units. Only AE hit records are retained. Records with an out-of-range channel, missing values, non-positive amplitudes or physically impossible feature combinations are removed. Raw values are converted to physical units per channel. The resulting table contains one row per hit, with frequency band and sensor location attached.

Saturation and censoring. Acquisition saturation and censoring are retained as data quality information rather than removed silently. Removing them would bias amplitude and energy statistics towards lower values. Where an acquisition amplitude ceiling $A_{\max}$ is available, hits with $A \geq A_{\max}$ are flagged as clipped. The clipped fraction is reported per frequency band.

Duration is similarly flagged where an acquisition limit $D_{\max}$ is known. No universal duration cap is assumed. Where a repeated upper value suggests a dataset-specific acquisition limit, this is addressed in the corresponding dataset analysis. Saturated or censored hits are excluded only when their values would distort a specific result.

Amplitude and signal-to-noise screening. All hits have already passed the acquisition threshold, which provides the primary noise rejection. Where a secondary screen is applied, its form depends on the dataset.

For waveform-based screening, the signal-to-noise ratio compares the peak of the complete stored record with the peak of a leading pre-trigger interval used to estimate local noise,

$$\mathrm{SNR}_{\mathrm{lin}} = \frac{A_{\mathrm{peak,signal}}}{A_{\mathrm{peak,noise}}}, \qquad \mathrm{SNR}_{\mathrm{dB}} = 20 \log_{10}\left(\frac{A_{\mathrm{peak,signal}}}{A_{\mathrm{peak,noise}}}\right), \tag{3.1}$$

with the retention threshold and noise window stated per dataset. In other datasets, the secondary screen uses an absolute recorded amplitude criterion rather than a signal-to-noise measure.

The screening result is stored as a per-hit flag and is not applied destructively. Results can therefore be compared between the full and screened populations. This is important because screening changes the population being analysed and therefore also changes the meaning of hit-rate.

Operational cleaning and synchronisation. Where operational data are available, duplicate timestamps are collapsed and out-of-range values are removed. Slowly varying operational signals may be smoothed with a moving mean of width M,

$$\bar{x}_M(t) = \frac{1}{M} \sum_{i=0}^{M-1} x(t-i). \tag{3.2}$$

Smoothing is applied only to the operational record, not to AE waveforms or features.

The AE and operational systems use independent clocks. They are aligned using a known operational event, such as a load ramp or commanded speed change, and the resulting offset is verified per dataset. Operating state is then attached to each hit by interpolation. Datasets without operational measurements omit this stage.

Segmentation into analysis windows. Long recordings can span several operating states, so hits are grouped according to the nominal operating state defined by the experimental record. Boundaries are obtained from operator markers, commanded-state logs or measured motion, depending on the dataset. The measured operating variables are used to verify the assigned state rather than to define universal window bounds.

The aggregation level follows the experimental design. Where load is stepped within a recording, observations are grouped by load (or pressure) step. Where load is fixed for a recording and speed changes within it, observations are grouped by speed step. For the reciprocating datasets used in the multivariate analysis, complete mechanical cycles are additionally grouped into fixed-length cycle windows. Where several operating variables are combined, observations are grouped by their nominal operating-state combination within each frequency band.

Treatment of transitions depends on the available timing information. Where window boundaries are defined from measured motion or explicit state markers, transitions are excluded by construction. Where only commanded-state times are available, the interval between successive markers may also contain the approach to the commanded state. These dataset-specific limitations are described in the corresponding chapters.

Per-hit scalar features are summarised by the median within the relevant aggregation unit, while the measured operating variables are summarised over the same interval. Activity is represented by hit count or an exposure-normalised hit-rate. Aggregation is kept separate by frequency band. Sensor positions within a band are pooled where supported by the dataset. Spectral descriptors are aggregated at the level supported by each dataset and are therefore not necessarily defined on the mechanical-cycle windows. Dataset-specific choices are stated in the corresponding chapters.

3.4. Feature Extraction

Scalar and spectral descriptors are first defined per hit or stored transient and then aggregated within the analysis window defined for each dataset (section 3.3). These aggregated descriptors are used as response variables during operational characterisation and as inputs to the multivariate analysis.

The median is used because the per-hit distributions are generally skewed and the experimental structures differ between campaigns. Both scalar and spectral descriptors are included because operating conditions can affect amplitude as well as frequency content. The features below are computed over the complete transient. No additional framing or window function is applied and the Fourier transform is unwindowed, because the stored records already represent finite, threshold-triggered AE transients and the analysis uses the complete recorded waveform, including the pre-trigger segment.

Scalar features. Two groups of scalar features are used. The first consists of parameters recorded directly by the acquisition system for each threshold-triggered event: peak amplitude A, rise time,

duration D, ring-down counts C and acquisition energy E_{acq}. Rise time is the interval from the first threshold crossing to the peak. Duration is the interval from the first to the last threshold crossing, while ring-down counts record the number of threshold crossings within the event. These quantities therefore depend on both the transient and the detection threshold. The acquisition system additionally reports $\mathrm{RMS}_{\mathrm{acq}}$, which describes the background preceding the hit rather than the event itself and is defined below.

The second group is recomputed from the stored waveform. For a transient y of length N,

$$x_{\mathrm{RMS}} = \sqrt{\frac{1}{N}\sum_{n=1}^{N} y^2(n)}, \qquad x_{\mathrm{energy}} = \sum_{n=1}^{N} y^2(n), \tag{3.3}$$

and, with $\bar{y}$ the transient mean, the crest factor and bias-corrected excess kurtosis are

$$x_{\mathrm{crest}} = \frac{\max_n |y(n)|}{x_{\mathrm{RMS}}}, \tag{3.4}$$

$$x_{\mathrm{kurtosis}} = \frac{N-1}{(N-2)(N-3)}\left[(N+1)g_2 + 6\right], \qquad g_2 = \frac{\frac{1}{N}\sum_{n=1}^{N}(y(n)-\bar{y})^4}{\left[\frac{1}{N}\sum_{n=1}^{N}(y(n)-\bar{y})^2\right]^2} - 3. \tag{3.5}$$

Here g_2 is the biased excess kurtosis, which is zero for a Gaussian distribution. The prefactor applies the standard sample-size bias correction, giving the unbiased estimator of excess kurtosis commonly denoted G_2. Amplitude measures describe the overall level, while crest factor and kurtosis describe waveform shape and impulsiveness.

Waveform energy is not used as a separate analysis feature beyond the acquisition quantity E_{acq}. The quantity x_{energy} in equation (3.3) is used only to construct the derived ratio below. Acquisition RMS and waveform RMS are not the same quantity and are kept separate. The Vallen acquisition RMS, $\mathrm{RMS}_{\mathrm{acq}}$, is the true RMS of the background noise preceding the hit and excludes the hit signal itself [50]. It therefore describes the measurement background preceding detection rather than the detected transient. Waveform RMS, x_{RMS}, is recomputed over all samples of the stored transient, including the pre-trigger segment.

Two derived ratios separate event length from signal intensity. With duration D, ring-down counts C, and waveform energy x_{energy},

$$r_{\mathrm{D/C}} = \frac{D}{C}, \qquad r_{\mathrm{E/D}} = \frac{x_{\mathrm{energy}}}{D}, \tag{3.6}$$

where $r_{\mathrm{D/C}}$ is the mean interval between threshold crossings and $r_{\mathrm{E/D}}$ represents the mean power of the event. Hit count is retained where available.

Spectral features. Each stored transient is transformed using the unwindowed FFT,

$$Y(m) = \sum_{n=0}^{N-1} y(n)\exp\!\left(-i2\pi\frac{mn}{N}\right), \qquad A(m) = |Y(m)|, \tag{3.7}$$

so that $A(m)^2$ is the power in bin m. The derived features are peak frequency, power-weighted spectral centroid, $95\,\%$ spectral roll-off, and partial powers in fixed frequency ranges. Band energy and the spectral centroid are

$$E_{f_1:f_2} = \sum_{m=m_1}^{m_2} \left(A(m)\right)^2, \qquad f_c = \frac{\sum_{m=1}^{W} f_m A(m)^2}{\sum_{m=1}^{W} A(m)^2}. \tag{3.8}$$

The $95\,\%$ spectral roll-off frequency is $f_{95} = f_{m_{95}}$, where m_{95} is the smallest bin satisfying

$$\sum_{m=1}^{m_{95}} A(m)^2 \geq 0.95 \sum_{m=1}^{W} A(m)^2.$$

The partial power in a frequency range is

$$P_{f_1:f_2} = \frac{\sum_{m=m_1}^{m_2} A(m)^2}{\sum_{m=1}^{W} A(m)^2}. \tag{3.9}$$

The centroid, roll-off, and partial power therefore use the same power weighting.

The frequency ranges used for partial powers are stated in the relevant dataset chapter and are not identical between datasets. They are treated as within-dataset descriptors rather than absolute quantities for comparison between rigs. Together, these features capture changes in spectral shape that scalar amplitude features cannot describe.

The output of this stage is an aggregated feature table for each analysis. Its rows represent the relevant load steps, speed steps or mechanical-cycle windows and contain the features and operating variables required for that analysis. This table is used in the following stages.

3.5. Operational Characterisation

Each aggregated feature is characterised against the measured operating variables before correction is considered. This stage establishes which relationships exist and in which direction. It does not establish whether they are predictable.

Association is quantified by the Spearman rank correlation. For n paired observations of a feature x and an operating variable z, with $R(\cdot)$ the rank within the corresponding sample,

$$r_s = \frac{\sum_{i=1}^{n} \left(R(x_i) - \overline{R(x)}\right)\left(R(z_i) - \overline{R(z)}\right)}{\sqrt{\sum_{i=1}^{n} \left(R(x_i) - \overline{R(x)}\right)^2 \sum_{i=1}^{n} \left(R(z_i) - \overline{R(z)}\right)^2}}. \tag{3.10}$$

A rank-based measure is used because the campaigns contain few operating levels, because several AE features are strongly skewed, and because no functional form of the operational response is assumed in advance. It detects monotonic association only, so a response that changes direction within the tested range can yield a weak coefficient despite a strong dependence.

Coefficients are computed per repeat of the varied operating axis and then averaged. The repeat unit and the number of operating levels follow from the experimental design and are stated in each dataset chapter. A feature and band combination is screened as operationally associated when $|\bar{r}_s| \geq 0.7$. The number of repeats meeting this criterion is reported alongside the mean, because a pooled mean can conceal sign reversals between repeats. The threshold is a study-specific screening criterion rather than a significance level, and the coefficient describes the retained hit population rather than the feature in general. Features passing the screen are assessed for correction eligibility in section 3.6. Features failing it are retained as descriptive features.

3.6. Correction Eligibility

Features are first screened for operational association as described in section 3.5. Features passing that screen are then assessed for correction eligibility.

A feature is corrected only when its operational dependence can be modelled reliably. Four criteria are used: stability, repeatability, range coverage, and predictive adequacy. Stability, repeatability, and range coverage identify candidates for correction, while predictive adequacy determines whether the fitted correction is accepted.

The first criterion is stability. The direction of the relationship should remain consistent across repeats rather than reversing between them. From the K repeat-specific Spearman correlations $r_1, \ldots, r_K$, with mean

$$\bar{r}_s = \frac{1}{K} \sum_{k=1}^{K} r_k, \tag{3.11}$$

the sign-consistency fraction is

$$\phi = \frac{1}{K} \sum_{k=1}^{K} \mathbf{1} \left[\operatorname{sign}(r_k) = \operatorname{sign}(\bar{r}_s) \right], \tag{3.12}$$

where $\mathbf{1}(\cdot)$ is the indicator function. A value of $\phi = 1$ indicates that all repeat correlations have the same direction as the mean relationship. The spread of the repeat correlations is quantified by

$$s_r = \sqrt{\frac{1}{K-1} \sum_{k=1}^{K} (r_k - \bar{r}_s)^2}, \tag{3.13}$$

which is the sample standard deviation of the repeat-specific correlation coefficients.

The two quantities are interpreted together. A value of ϕ close to one together with a small s_r indicates a relationship whose direction and strength are consistent across repeats, whereas lower sign consistency or a large spread indicates instability. No fixed numerical threshold is imposed on either quantity; they are used as descriptive evidence alongside the remaining eligibility criteria.

The second criterion is repeatability. Where the same commanded state is visited more than once, the feature should return a reproducible value. Repeatability is quantified using the symmetric relative difference

$$d = 100 \, \frac{|v_a - v_b|}{\frac{1}{2}\left(|v_a| + |v_b|\right)}, \tag{3.14}$$

where v_a and v_b are the feature values at two visits. A feature is treated as repeatable when its median difference is below $10\,\%$, marginal between $10\,\%$ to $25\,\%$, and unreliable above $25\,\%$. These are study-specific screening thresholds rather than universal AE limits. The maximum difference is also reported because sparsely populated operating-state combinations can produce unstable feature estimates and therefore large individual differences.

The third criterion is range coverage. A correction is estimated only over an operating range that is adequately populated and is not extrapolated beyond that range.

A feature meeting these three criteria becomes a correction candidate. It is accepted only when the fitted model also provides adequate prediction on reference observations that were not used for estimation. Features that fail any criterion remain in the analysis as descriptive features, but are not treated as corrected indicators.

The complete sequence can only be applied when the dataset contains a suitable reference condition, adequate operational coverage, and an independent validation structure. Where a criterion cannot be tested, it is reported as not testable rather than assumed to be satisfied.

3.7. Feature-Specific Correction

Operational correction models are estimated from reference data where a suitable reference condition is available. This follows SHM normalisation approaches in which measured environmental or operational variables are used to model expected reference behaviour [48, 13, 47]. Condition labels are not used to estimate the operational model. They are introduced after correction when the remaining feature structure is interpreted.

For a feature x_j, the expected value under the measured operating state is estimated as

$$\hat{x}_j = f_j(\mathbf{z}), \qquad r_j = x_j - \hat{x}_j, \tag{3.15}$$

where $\mathbf{z}$ contains the measured operating variables and r_j is the residual. The residual therefore represents variation not explained by the fitted reference operating response [47].

The operational model is kept deliberately simple and is fitted only over the range supported by the reference data. The form of the model is specified for each dataset according to its experimental design and available operating coverage.

Correction effectiveness is assessed by examining whether residual association with the measured operating variables is reduced and by inspecting the residuals for remaining systematic operating structure. The raw features are retained so that raw and residual representations can be compared.

3.8. Multivariate Residual Analysis

The correction stage treats features individually. The multivariate analysis then examines whether combinations of features retain condition-associated structure after measured operational effects have been reduced. Because the relevant operating variables are measured, their influence is reduced explicitly before the multivariate analysis rather than being inferred from PCA alone [13, 14, 47].

PCA is used as the linear multivariate diagnostic. PCA is widely used in SHM for pattern evaluation and feature visualisation [14], and has also been applied to multivariate AE features in low-speed bearing fault diagnosis [41]. In chapter 6, TB3 instead uses baseline-fitted residualisation as an exploratory sensitivity analysis because the same independent validation structure is not available.

Features are transformed and scaled where required by the subsequent analysis. Scaling parameters are estimated once and then applied unchanged to all observations in that representation. The reference population used for each representation is stated in chapter 6.

PCA is interpreted through explained variance, component loadings and the association of component scores with operating variables and condition where available. Reporting the loadings keeps the interpretation linked to the original physical features. The representation is treated as diagnostic rather than as a classifier because PCA follows directions of maximum variance and does not guarantee that the most useful features for discrimination are preserved [14]. Condition labels are used only to interpret the resulting representation and not to construct it.

The main comparison is between the raw and corrected feature spaces within a single dataset. Both are examined with respect to operating state and condition. If organisation by operating state weakens after correction while condition-associated structure remains, the two sources of structure are considered distinguishable within that dataset.

3.9. Evaluation

The workflow is evaluated as an operationally informed analysis rather than as a validated diagnostic monitor.

Operational correction is assessed by whether association with the measured operating variables is reduced and whether systematic operating structure remains in the residual representation. Multivariate analysis then examines whether organisation by operating state is reduced while condition-associated structure remains.

Transferability is assessed through application of the same characterisation and eligibility logic across datasets. Numerical feature relationships are not assumed to transfer between rigs. Because several analyses use few operating levels and correlated observations, the quantitative results are interpreted as trend evidence and formal statistical significance is not claimed.

3.10. Software Implementation

The workflow is implemented in Python using reusable modules for data loading, AE and operational cleaning, synchronisation, feature extraction, operating state definition, correction and embedding. The implementation uses `NumPy` [51], `pandas` [52], `SciPy` [53], `statsmodels` [54] and `scikit-learn` [55]. This structure allows the same workflow to be applied consistently across datasets.

4

Operational Characterisation of AE Features in Linear Bearing Datasets

This chapter characterises how load, stroke speed and lubrication or contamination state influence AE features in the linear bearing datasets. It addresses the first research subquestion and provides supporting evidence for the second by distinguishing supported, unstable and untestable operational relationships. The full correction and validation procedure is reserved for Chapter 5.

The analysis covers the Phase 4 advanced-degradation dataset and two Phase 2 blocks using initially pristine mechanical components: an intact baseline (TB2) and a lubrication-contamination progression (TB3). The central criterion is stability. Operational association is screened first, but a relationship is treated as predictable only if its direction and form persist across the variation available in the experiment. Repeatability and range coverage are then assessed where the design permits, while SNR sensitivity is treated separately as processing robustness.

4.1. Chapter Scope and Shared Analysis Strategy

The campaigns use the same aggregation and Spearman-correlation framework, but their clean operating axes differ. Phase 4 varies load within a recording, while speed and lubrication vary between recordings and are therefore exposed to session and re-coupling effects. Phase 2 reverses this structure: speed varies within recordings and load between them; TB3 additionally varies contamination between runs. TB2's load axis is withdrawn because the detection threshold changes with load (section 4.3.3).

The repeat unit and correlation resolution also differ. Phase 4 load correlations use eighteen tests and seven load levels. Phase 2 load correlations span four loads, while speed correlations span only three states and can take only ± 1 or ± 0.5. Phase 2 speed coefficients are therefore interpreted as rank-order indicators. Its ascending–descending speed programme also provides the within-session repeatability check used in section 4.3.5. Table 4.1 summarises these design differences and the strongest evidence available from each dataset.

The Chapter 3 criteria are applied consistently: Spearman correlation screens association; directional agreement tests stability; repeated commanded states test repeatability; and sampled operating extent determines range coverage. SNR sensitivity is reported separately because it changes the retained hit population rather than the physical state. Predictive adequacy is not a Chapter 4 verdict because these datasets do not support the complete correction-validation design. Each dataset therefore ends with an evidence assessment, with unavailable criteria reported as not testable and criteria not evaluated reported as not performed. The common sequence can be found in section B.2.

For load sensitivity, hits are assigned to load steps and aggregated by step and frequency band. Hit-rate and median per-hit features are retained, together with duration/counts and energy/duration. Spearman correlation with measured force F_z is then calculated for each test; $|r_s| \geq 0.7$ is used as a strong

Dataset	Clean operating axis	Resolution	Additional evidence	Main limitation
Phase 4	Load	Seven load levels	Eighteen load sweep tests permit assessment of directional stability across tests	Speed and lubrication vary between recordings. Direct revisit repeatability is not available
Phase 2 TB2	Stroke speed	Three speed levels	Ascending and descending visits permit direct within session repeatability assessment	Load varies between recordings and is additionally confounded with the acquisition threshold
Phase 2 TB3	Stroke speed	Three speed levels	Operational relationships can be compared across successive contamination states	Load varies between recordings and contamination progresses sequentially between runs

Table 4.1: Experimental structure and analytical role of the linear bearing datasets.

effect-size screen rather than a significance test [56, 57]. Speed and lubrication are evaluated through matched comparisons, with magnitudes normalised within matched groups where required. Both the acquisition-supplied features and descriptors recomputed from the stored waveforms are used.

Pipeline instantiation. The five preprocessing stages of section 3.3 are applied in the same order, with dataset-specific choices summarised in table 4.2; segmentation follows the experimental design: load step for Phase 4, recording plus speed marker for TB2, and per-hit marker reconstruction for TB3.

4.2. Phase 4: Advanced Degradation State

The Phase 4 Linear Bearing dataset was acquired in the Huisman Equipment R&D laboratory [58]. Its within-recording load sweep provides the chapter's cleanest load axis. Stroke speed and lubrication are treated as secondary between-recording modifiers, while SNR filtering is used only as a robustness check.

4.2.1. Recorded Data Overview

The campaign contains approximately 7.9 million hits across twenty-one acquisition files. Activity spans all three frequency bands and multiple sensor locations; the busiest channel contributes about one fifth of the total. This supports band-specific analysis without reducing the dataset to a single sensor position. Population distributions are reported in section B.3.1.

4.2.2. Bearing Condition and Experimental Programme

The rig uses a double linear bearing around a central sliding element, with internal and external raceways. The bearing was at the end of its service life and inspection confirmed severe degradation [58]; no single dominant failure mode is identified. Results are therefore interpreted as those of an advanced degradation state rather than a localised defect.

Phase 4 was recorded in sessions of three consecutive tests, referred to here as triplets. Within a triplet the lubrication condition and the analysed sensor node are fixed, and the three tests are run at the three stroke speeds ($0.095\,\mathrm{m\,s^{-1}}$, $0.140\,\mathrm{m\,s^{-1}}$ and $0.165\,\mathrm{m\,s^{-1}}$), with load swept within each test. The lubrication condition is changed between triplets. Three conditions were imposed: starvation (group 1), full-mixed (group 2) and full-flushed (group 3). Each condition was recorded in two triplets, one analysed at a raceway node and one at a traveller node, giving the eighteen load-sweep tests of the primary analysis. Tests are labelled condition–triplet–speed, so test 3-2-1 is the lowest-speed test of the second full-flushed triplet. Load therefore varies within each test, speed between the tests of a triplet, and lubrication between triplets. An additional full-flushed triplet, group 3-3, was recorded at approximately constant load to isolate the speed response (section 4.2.5). Per-hit features and stored waveforms are taken from the acquisition system, and force, velocity and cycle count from SCADA.

Stage	Phase 4	TestBlock 2	TestBlock 3
Validation and units	Hit records with a valid channel; numeric coercion; A_{dB} formed only where $A > 0$. No records removed	Hit records only. No null channels and no non-positive amplitudes are present, so no records are removed	As TestBlock 2
Saturation and censoring	Amplitude clipping flagged at $A_{\max} = 49\,990\,\mu\mathrm{V}$; clipped fraction reported per band	Same ceiling. 644 of $1\,909\,077$ hits clipped ($0.034\,\%$), distributed without a consistent directional pattern across the four load steps	Same ceilings. Low-band duration censoring is treated as bounded here and excluded from the primary vector in chapter 6
Amplitude and SNR screening	Waveform signal to noise ratio against a $100\,\mu\mathrm{s}$ pre-trigger noise window; $\mathrm{SNR} \geq 5$ and ≥ 10 reported as a sensitivity check	Same definition, referred to the block's own acquisition floor; retained fractions $100\,\%$, $14.4\,\%$ and $0.45\,\%$	No SNR-margin filter in the primary analysis; retained-population sensitivity is checked separately using the low-band acquisition-floor threshold
Operational cleaning and synchronisation	SCADA at approximately $1\,\mathrm{Hz}$ providing F_z, v and stroke count; fixed clock offset verified per test against the force ramp; F_z smoothed with a centred moving mean	One continuously logged SCADA file spans the four AE recordings, so alignment is established per recording	Operating state reconstructed from operator markers; SCADA record checked separately
Segmentation into analysis units	Load step from the AE markers, confirmed against the SCADA force record; activity expressed per load cycle	Load is the recording; speed step taken from markers within each recording, with ascending and descending passes retained separately	Load and speed reconstructed per hit from the markers, because several files combine two or three load steps

Table 4.2: Instantiation of the preprocessing stages of section 3.3 in each linear bearing dataset.

4.2.3. Acquisition and Analysis Constraints

Each of the six sensor nodes carries a Physical Acoustics R6α, R15α and WSα transducer, with nominal responses of $60\,\mathrm{kHz}$ resonant, $150\,\mathrm{kHz}$ resonant and broadband respectively. The corresponding acquisition channels are band-pass filtered to $40\,\mathrm{kHz}$ to $100\,\mathrm{kHz}$, $95\,\mathrm{kHz}$ to $180\,\mathrm{kHz}$ and $180\,\mathrm{kHz}$ to $580\,\mathrm{kHz}$, referred to here as the low, mid and high bands. The band definitions are therefore acquisition settings rather than transducer operating ranges. The complete channel-to-sensor mapping is given in section B.1. Node positions are shown in figure 4.1: A–B are drive-end raceway locations, C–D are bearing-body locations, and E–F are far-end raceway locations.

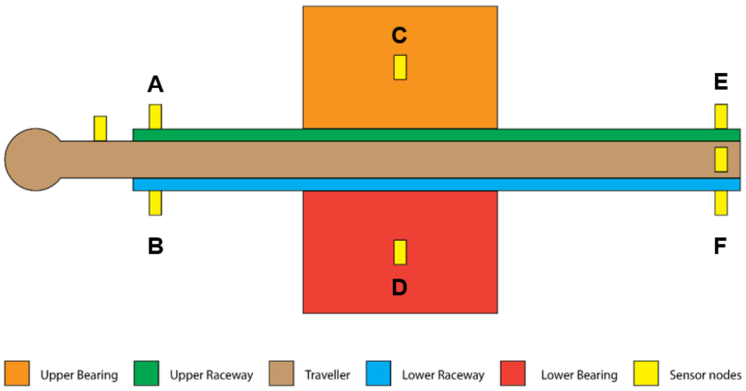


Figure 4.1: Sensor node locations (A–F) on the Phase 4 linear bearing rig.

Signals pass through $40\,\mathrm{dB}$ Vallen AEP5H pre-amplifiers to an AMSY-6 system. Trigger thresholds are $55\,\mathrm{dB}$ in the low band and $50\,\mathrm{dB}$ in the mid and high bands. Each trigger stores a $812\,\mu\mathrm{s}$ transient at $5\,\mathrm{MHz}$, including a $200\,\mu\mathrm{s}$ pre-trigger segment. Each stored waveform is linked to its corresponding hit record by the acquisition system. Amplitude is reported in $\mu\mathrm{V}$ or, where required, relative to $1\,\mu\mathrm{V}$ as

$$A_{\mathrm{dB}} = 20\,\log_{10}\!\left(\frac{A}{1\,\mu\mathrm{V}}\right). \tag{4.1}$$

Sensor-referred thresholds include the $40\,\mathrm{dB}$ pre-amplifier gain, whereas `.pridb` amplitudes are digitiser-referred.

The Phase 4 implementation is summarised in table 4.2. SNR uses the $100\,\mu\mathrm{s}$ pre-threshold window, with $\mathrm{SNR}_{\mathrm{lin}} \geq 5$ ($14\,\mathrm{dB}$) where filtering is applied [58]. SCADA provides stroke count, velocity v and force F_z at approximately $1\,\mathrm{Hz}$; a fixed offset is verified against the AE load-step markers. AE activity is normalised by load cycle rather than time [58]. With ten load cycles per machine stroke, $C_n = 10N_{\mathrm{stroke}}$, and the hit count for cycle i is

$$N_h(i) = \sum_{m=1}^{M_h} \mathbf{1}(i \leq C_{n,\mathrm{AE}}(m) < i+1)\,, \tag{4.2}$$

where M_h is the number of retained hits and $\mathbf{1}(\cdot)$ is the indicator function. Load-step segmentation follows the AE markers and is confirmed against SCADA force.

Acquisition ceilings. Amplitude is clipped at $A_{\max} = 49\,990\,\mu\mathrm{V}$ in $7.43\,\%$ of hits overall (low/mid/high: $5.73\,\%$, $11.84\,\%$, $8.04\,\%$). Duration reaches the $104\,857.6\,\mu\mathrm{s}$ counter limit in $5.29\,\%$ overall ($3.87\,\%$, $8.47\,\%$, $6.65\,\%$ by band). These ceilings bound amplitude- and duration-related quantities, particularly in the mid band. Reported medians remain below the limits, but load dependence of the censored fraction has not been established; the affected relationships are therefore qualified rather than corrected.

4.2.4. Influence of Load

Load is the primary Phase 4 axis. Two feature groups are evaluated: the quantities supplied by the acquisition system, and the scalar and spectral descriptors recomputed from the stored transients,

which give a complementary view of the same relationships. Measured force increases monotonically across load steps in all eighteen tests, confirming synchronisation.

Each stored waveform inherits its load step, band and F_z from the corresponding hit record. The recomputed group comprises four scalar and six spectral descriptors, aggregated by load step and band, with no additional SNR filter. Acquisition RMS and waveform RMS are not interchangeable: $\mathrm{RMS}_{\mathrm{acq}}$ is the background level preceding the threshold crossing, whereas x_{RMS} is recomputed over the full stored transient. Their load responses agree only in the high band, so comparisons are made per feature and band.

Table 4.3 reports the mean Spearman correlation with F_z. Parentheses give the number of tests satisfying $|r_s| \geq 0.7$, bold marks $|\bar{r}_s| \geq 0.7$, and hit-rate is common to both groups and reported once. Per-test matrices are given in sections B.4 and B.5.

Quantity	Low band	Mid band	High band
Acquisition quantities			
Hit-rate	$+0.45$ (8)	$+0.61$ (10)	$\mathbf{+0.89}$ (16)
RMS (acquisition)	-0.59 (11)	-0.14 (3)	$\mathbf{+0.84}$ (14)
Energy	$+0.63$ (9)	$\mathbf{+0.86}$ (14)	$+0.41$ (3)
Amplitude [dB]	$+0.59$ (8)	$\mathbf{+0.89}$ (17)	$\mathbf{+0.77}$ (12)
Counts	$+0.67$ (9)	$\mathbf{+0.83}$ (15)	-0.17 (3)
Duration	$+0.67$ (10)	$\mathbf{+0.75}$ (13)	-0.14 (3)
Rise time	$+0.41$ (6)	-0.60 (9)	-0.51 (8)
Duration / counts	-0.54 (10)	$\mathbf{-0.81}$ (15)	$+0.19$ (4)
Energy / duration	$+0.56$ (10)	$\mathbf{+0.86}$ (16)	$+0.69$ (10)
Waveform-derived quantities			
Waveform RMS	$+0.56$ (10)	$\mathbf{+0.92}$ (17)	$\mathbf{+0.77}$ (15)
Waveform peak	$+0.60$ (10)	$\mathbf{+0.91}$ (17)	$+0.69$ (17)
Crest factor	$+0.35$ (8)	$\mathbf{+0.70}$ (14)	$+0.49$ (13)
Kurtosis	$+0.40$ (8)	$\mathbf{+0.77}$ (15)	$+0.38$ (12)
Peak frequency	-0.05 (10)	$+0.36$ (12)	-0.41 (9)
Spectral centroid	$+0.57$ (11)	$+0.35$ (11)	-0.26 (11)
Spectral roll-off $95\,\%$	$+0.54$ (14)	$+0.69$ (12)	$+0.21$ (11)
Partial power $40\,\mathrm{kHz}$ to $100\,\mathrm{kHz}$	$\mathbf{-0.86}$ (15)	-0.15 (10)	-0.01 (12)
Partial power $100\,\mathrm{kHz}$ to $300\,\mathrm{kHz}$	$\mathbf{+0.82}$ (14)	$+0.16$ (10)	$+0.23$ (13)
Partial power $300\,\mathrm{kHz}$ to $1000\,\mathrm{kHz}$	$+0.62$ (8)	-0.34 (11)	-0.23 (13)

Table 4.3: Phase 4 load sensitivity, pooled across the eighteen load sweep tests.

Load sensitivity is strongly band dependent. High-band hit-rate is the most consistent activity measure; energy, amplitude and count-related features are strongest in the mid band. Acquisition RMS is positive in the high band but tends negative in the low band, and duration-related features concentrate in the low and mid bands. Among the recomputed quantities, waveform RMS and peak are strongly load positive in the mid and high bands, crest factor and kurtosis are weaker, and the single-value spectral descriptors combine moderate means with frequent directional changes. Each feature–band combination is therefore assessed separately. Figure 4.2 adds the stability measures s_r and ϕ from section 3.6. High-band hit-rate, mid-band waveform RMS and the low-band partial powers combine strong means with low spread and ϕ near one. In contrast, low-band peak frequency has $\phi = 0.50$, showing that its near-zero mean reflects directional inconsistency rather than absence of response.

The representative 3-2 full-flushed traveller-node triplet shows that low-band load coupling strengthens with stroke speed, whereas mid- and high-band coupling is already present. It also shows opposite load trends in the $40\,\mathrm{kHz}$ to $100\,\mathrm{kHz}$ and $100\,\mathrm{kHz}$ to $300\,\mathrm{kHz}$ partial powers ($r = -0.86$ and $+0.82$). This is reported as spectral redistribution only; the second range lies above the nominal low-band passband, so the physical mechanism is unresolved. The underlying matrices are given in section B.5.

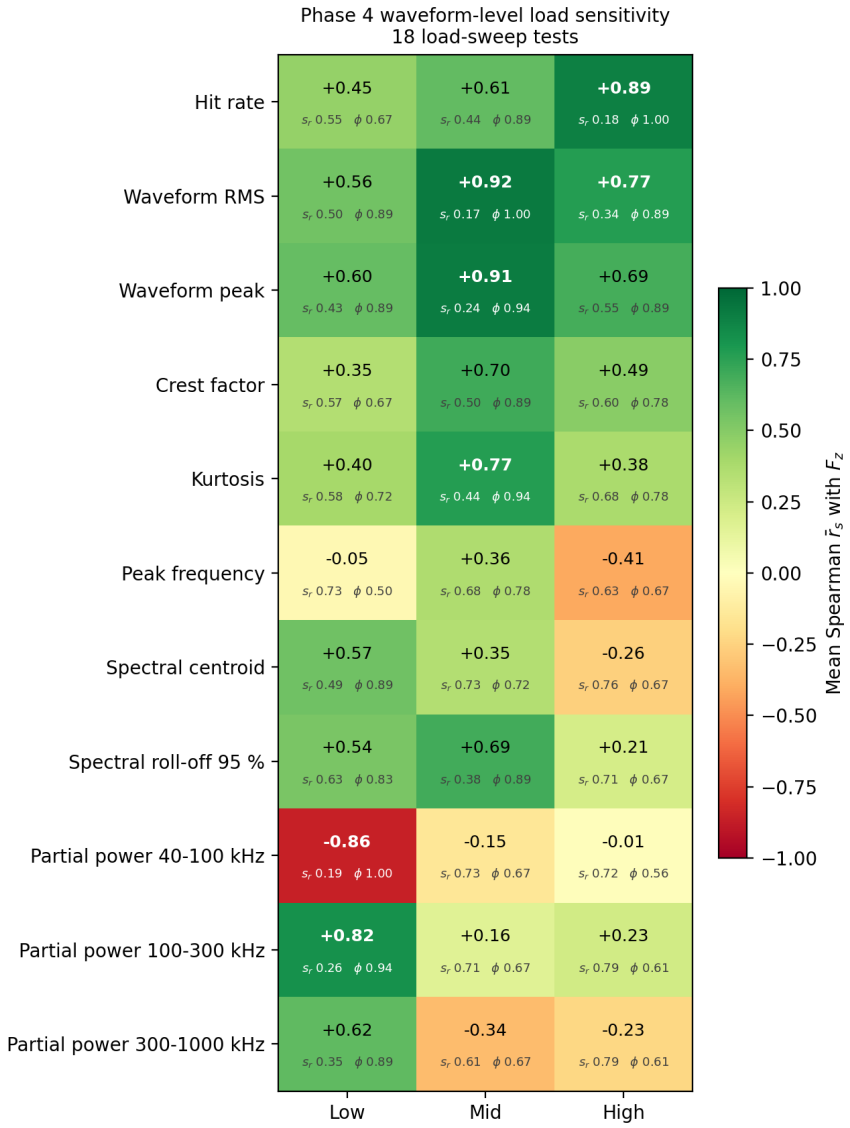


Figure 4.2: Phase 4 waveform level load sensitivity across the eighteen load sweep tests.

Sensitivity to retained hit population. The waveform analysis is repeated for all hits, $\mathrm{SNR} \geq 5$ and $\mathrm{SNR} \geq 10$. The unfiltered population remains primary because filtering changes both the amplitude distribution and the meaning of hit-rate. Waveform RMS and peak retain their direction and strength, whereas hit-rate, crest factor, kurtosis, spectral descriptors and low-band partial powers are population sensitive (section B.6). This is treated as processing robustness, not physical stability.

4.2.5. Influence of Stroke Speed

Stroke speed is a between-recording modifier in Phase 4. Median amplitude is comparatively stable, whereas hit-rate, energy, counts and duration-related features vary more strongly; 50 of 162 feature–band–condition combinations change direction (section B.9). Speed must therefore be controlled when comparing feature levels between tests.

The constant-load 3-3 triplet (approximately $497\,\mathrm{kN}$) isolates this effect (figure 4.3). Hit-rate is most speed sensitive: the high band roughly doubles, the mid band falls by a factor of six, and the low band changes direction. High-band RMS and peak decrease while hit-rate rises, showing that event rate and transient amplitude respond differently to speed. Spectral position descriptors remain comparatively stable.

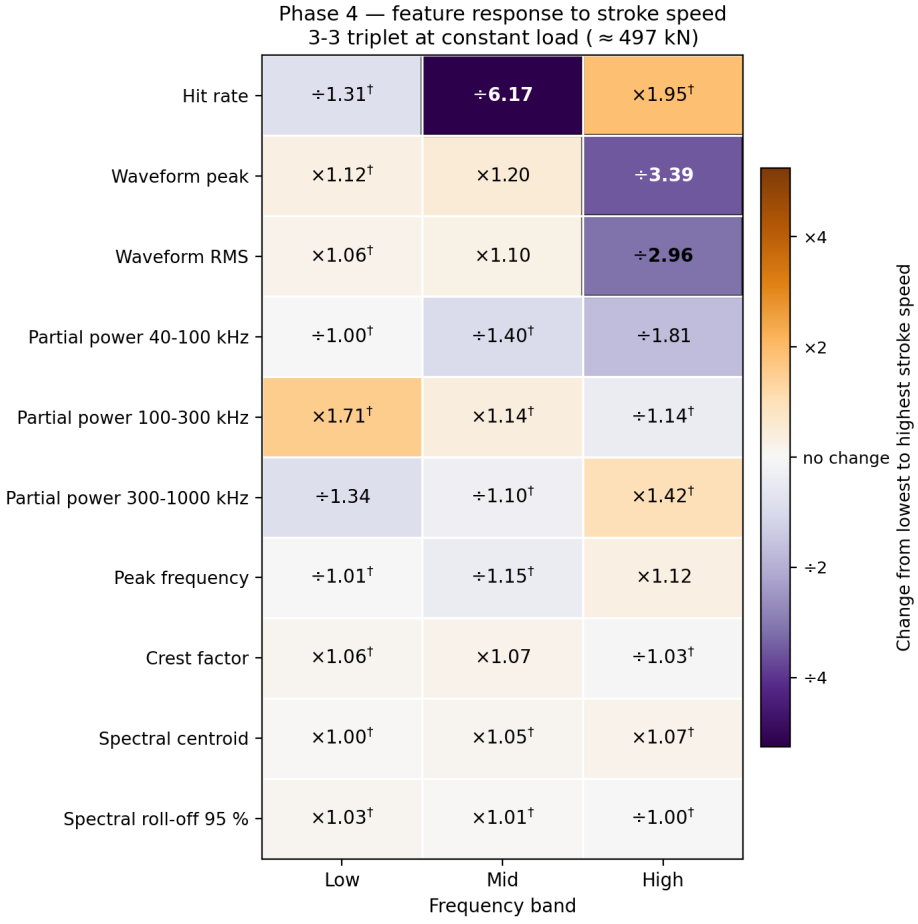


Figure 4.3: Phase 4 feature response to stroke speed in the 3-3 constant load triplet ($497\,\mathrm{kN}$).

4.2.6. Influence of Lubrication State

Lubrication is imposed rather than measured and is evaluated at matched speeds relative to lifetime grease. Effects are feature- and band-dependent: mid-band energy reaches four times the lifetime-grease level under full-mixed grease, while median amplitude remains within $10\,\%$ in all bands (sec-

tion B.10). Without measured film thickness or grease distribution, lubrication is treated as a descriptive modifier rather than a correction axis.

4.2.7. Assessment of Supported Operational Relationships

Table 4.4 assesses the Phase 4 load relationships on association and directional stability. Association marks $|\bar{r}_s| \geq 0.7$ in table 4.3. Stability for the acquisition quantities is judged from the per-test matrices in section B.4, while waveform-derived stability is summarised in figure 4.2. Range coverage is adequate over the tested load range. Repeatability cannot be assessed: the load sweep visits each commanded load once per test, so no feature value can be compared between two visits to the same state. It is omitted rather than recorded as a criterion met, and no Phase 4 relationship should be read as demonstrated repeatable. The table records supported relationships rather than accepted corrections.

Feature (band)	Association	Stability
Acquisition quantities		
Amplitude (mid)	✓	✓
Amplitude (high)	✓	✓
Energy (mid)	✓	✓
Energy / duration (mid)	✓	✓
Counts (mid)	✓	✓
Duration (mid)	✓	✓
Duration / counts (mid)	✓	✓
RMS (acquisition, high)	✓	✓
Hit-rate (high)†	✓	✓
Waveform-derived quantities		
Waveform RMS (mid)	✓	✓
Waveform RMS (high)	✓	✓
Waveform peak (mid)	✓	✓
Kurtosis (mid)†	✓	(✓)
Crest factor (mid)†	✓	(✓)
Partial power $40\,\mathrm{kHz}$ to $100\,\mathrm{kHz}$ (low)†	✓	✓
Partial power $100\,\mathrm{kHz}$ to $300\,\mathrm{kHz}$ (low)†	✓	✓

Table 4.4: Phase 4 assessment of load relationships. A parenthesised tick denotes less consistent direction across the eighteen sweeps; † marks features sensitive to the retained hit population. Repeatability is not testable and is omitted.

Supported Phase 4 load relationships are concentrated in the amplitude anchors, waveform RMS and peak, and the mid-band energy-, count- and duration-related quantities. The term amplitude anchors refers here to the median hit amplitude in the mid and high bands. Median amplitude is used as an anchor because it retains a consistent operational relationship in all three linear bearing datasets (section 4.5.1) and therefore serves as a reference against which the less stable features are compared. High-band hit-rate is also associated and directionally stable, but its interpretation remains sensitive to the retained hit population. Waveform RMS and peak retain their response under SNR filtering, whereas mid-band kurtosis and crest factor are both less directionally consistent and more population sensitive.

The two low-band partial-power descriptors show strong load relationships with consistent direction across the sweeps, but their magnitude depends on stroke speed and their response changes under SNR filtering. They are therefore retained as qualified operational relationships. Peak frequency, spectral centroid and spectral roll-off show weaker pooled association and directional changes between tests. Unsupported or qualified features remain descriptive rather than being discarded. Complete correction and validation are addressed in Chapter 5.

4.3. Phase 2 TestBlock 2: Intact Bearing

The Phase 2 campaign used a modular double linear bearing rig in the Huisman Equipment R&D laboratory [33]. This thesis uses the two available blocks: the intact baseline, TestBlock 2 (TB2), and the lubrication-contamination block, TestBlock 3 (TB3; section 4.4). TB2 provides the clean-bearing operational reference; TB3 tests how those relationships change with lubricant condition.

Load step	Load	Cycles per speed		
		Speed 0	Speed 1	Speed 2
Load 0	$100\,\mathrm{kN}$ ($6\,\%$ max)	20	40	50
Load 1	$260\,\mathrm{kN}$ ($20\,\%$ max)	20	40	50
Load 2	$785\,\mathrm{kN}$ ($50\,\%$ max)	10	20	30
Load 3	$1305\,\mathrm{kN}$ ($90\,\%$ max)	10	20	30

Table 4.5: Phase 2 baseline load and excitation matrix.

4.3.1. Recorded Data Overview

TB2 contains approximately 1.9 million hits across four files. More than $99\,\%$ of activity lies in the low band and is concentrated on two lower-chamber support channels. TB2 is therefore treated as a low-band, effectively single-location dataset; mid- and high-band load–speed combinations are too sparse for stable medians. Population details are given in section B.3.2.

4.3.2. Experimental Programme and Data Quality

TB2 and TB3 use pristine Hardox 600 raceways and $75\,\mathrm{mm}$ rollers, giving identical nominal geometry. The rig applies vertical load through two static support rings, between which a nose ring translates horizontally to generate the reciprocating motion [33]. The nose ring acoustically separates an upper and a lower chamber, and sensor locations are named by chamber, by mounting on the raceway or the substructure, and by whether they sit on the support ring or the nose ring.

Speeds 0, 1 and 2 correspond to $24\,\mathrm{s}$, $18\,\mathrm{s}$ and $12\,\mathrm{s}$ per $140\,\mathrm{mm}$ stroke, respectively.

TB2 was recorded on 2 October 2020 in the as-built undamaged state. Each load is stored in a separate file, with speed ramped $0 \to 1 \to 2 \to 1 \to 0$ within the file. The repeated lower speeds provide the repeatability check of section 4.3.5. Of $1\,909\,077$ hits, $1\,871\,014$ have linked waveforms.

The acquisition hardware matches Phase 4 but uses eight sensor locations and 24 channels (table B.2). Channels 1–12 are in the lower chamber and 13–24 in the upper. The nominal threshold is $50\,\mathrm{dB}$, but channels 1 and 7 were retuned to $56\,\mathrm{dB}$ and then $62\,\mathrm{dB}$; this directly affects the load analysis. Other acquisition settings follow Phase 4.

Acquisition ceilings. Amplitude clipping affects $0.034\,\%$ of TB2 overall and only $0.014\,\%$ of the low band, so it is negligible for the reported low-band analysis. Duration reaches the $104\,857.6\,\mu\mathrm{s}$ counter limit in $1.40\,\%$ of low-band hits; duration-based repeatability results are therefore interpreted as bounded quantities.

Band concentration. The retained population here comprises hits assigned to a marked load–speed combination and frequency band after speed segmentation, and therefore differs from the $1\,909\,077$ database-hit count, which also includes hits recorded outside the marked speed steps. Of the $1\,889\,865$ retained hits, $1\,886\,543$ are in the low band, compared with 2832 mid-band and 490 high-band hits. The low band therefore contains $99.82\,\%$ of retained activity. Sparse mid- and high-band load–speed combinations are omitted rather than summarised by unstable medians.

This restriction is specific to TB2; TB3 retains approximately $145\,000$ mid-band and $61\,000$ high-band hits and is analysed across all three bands.

Channel concentration and threshold retuning. Channels 1 and 7 contain $1\,893\,021$ of $1\,909\,077$ hits ($99.16\,\%$) and dominate the low-band analysis. Their threshold was raised from $50\,\mathrm{dB}$ to $56\,\mathrm{dB}$ and then $62\,\mathrm{dB}$. The minimum amplitudes in table 4.6 track this change across load recordings, so threshold and load increase together.

TB3 does not show this fault: channels 1 and 7 remain between $50.89\,\mathrm{dB}$ and $51.38\,\mathrm{dB}$ across all sixteen run–load combinations, despite carrying $90.56\,\%$ of hits.

Load step	Load	Min. A_{dB}, ch. 1	Min. A_{dB}, ch. 7
Load 0	$100\,\mathrm{kN}$	50.90	51.24
Load 1	$260\,\mathrm{kN}$	56.89	56.96
Load 2	$785\,\mathrm{kN}$	57.74	57.86
Load 3	$1305\,\mathrm{kN}$	63.07	62.86

Table 4.6: Minimum recorded hit amplitude per TestBlock 2 load step on channels 1 and 7.

4.3.3. Withdrawal of the Load Axis

The threshold rises by approximately $12\,\mathrm{dB}$ with the load sequence. Such retuning alone raises retained amplitude measures, shortens threshold-defined duration and counts, and reduces detected activity. The observed TB2 load correlations follow this pattern: amplitude $+0.93$, waveform RMS $+1.00$, duration -0.80 and counts -0.67, while hit-rate spans 18.9–$17\,410$ hits per minute without load ordering. Acquisition RMS also reads $+1.00$; because it measures the background preceding each retained hit, this reflects the changing retained population rather than the truncation itself.

Because no constant-load control or unretuned high-activity channel separates threshold from load, the TB2 load axis is withdrawn. The correlations are retained only as a record of the confound in section B.7. TB2 therefore contributes only the speed and repeatability analyses; the Phase 2 load evidence comes from TB3.

4.3.4. Influence of Stroke Speed

Speed is TB2's clearest operating axis, varying within each recording at a fixed threshold, with load as the repeat unit. With only three speed states, the Spearman coefficients are highly discretised (± 1 or ± 0.5) and are therefore interpreted primarily as rank ordering.

Table 4.7 reports the low-band results for the unfiltered, $\mathrm{SNR} \geq 5$ and $\mathrm{SNR} \geq 10$ populations, referenced to the TB2 acquisition floor of $51\,\mathrm{dB}$. Parentheses give the number of load repeats satisfying $|r_s| \geq 0.7$, and bold marks $|\bar{r}_s| \geq 0.7$.

In the unfiltered population, energy, median amplitude, duration, rise time and waveform peak are strictly ordered with speed across all four loads; waveform RMS and crest factor are ordered in three of four. Hit-rate and the single-value spectral descriptors are weaker. SNR filtering substantially changes these relationships while retaining only $0.45\,\%$ of low-band hits at $\mathrm{SNR} \geq 10$: the mean speed correlation for median amplitude changes from $+1.00$ to $+0.50$ and then -0.62, the waveform-RMS correlation reverses from $+0.88$ to -1.00, and the hit-rate correlation strengthens to $+1.00$. Because the filter is amplitude-referenced and heavily alters the population, the unfiltered result remains primary and the filtered cases are treated as sensitivity checks.

4.3.5. Ascending and Descending Repeatability

Speeds 0 and 1 are revisited on the descending ramp, providing the chapter's direct within-session repeatability test. Low-band feature medians are compared between passes using the symmetric relative difference. Speed 2 is visited once, and one empty Load 0/Speed 1 pass is excluded, leaving seven comparable load–speed combinations.

Table 4.8 reports the results for both feature groups. Median differences below $10\,\%$ are classified as repeatable, $10\,\%$ to $25\,\%$ as marginal and above $25\,\%$ as unreliable. Rows are ordered by median difference within each group.

The repeatability check separates association from reproducibility. Rise time is perfectly rank ordered with speed but differs by about $46\,\%$ between repeated states and is therefore unreliable. Most waveform-derived descriptors are repeatable, with several spectral descriptors within $1\,\%$; median amplitude and waveform RMS are also repeatable. Counts and rise time are unreliable, while duration and energy are marginal. The sparsely populated Load 3/Speed 0 combination drives several extreme maxima, including acquisition RMS at $171.8\,\%$ and kurtosis at $200\,\%$, but does not change the median classifications.

Quantity	No filter	$\mathrm{SNR} \geq 5$	$\mathrm{SNR} \geq 10$
Acquisition quantities			
Hit-rate	$+0.62$ (3/4)	$\mathbf{+1.00}$ (4/4)	$\mathbf{+1.00}$ (4/4)
RMS (acquisition)	$+0.38$ (3/4)	-0.25 (2/4)	$+0.25$ (2/4)
Energy	$\mathbf{+1.00}$ (4/4)	$+0.38$ (1/4)	-0.62 (3/4)
Amplitude [dB]	$\mathbf{+1.00}$ (4/4)	$+0.50$ (2/4)	-0.62 (3/4)
Counts	$+0.62$ (3/4)	$+0.28$ (2/4)	$+0.62$ (3/4)
Duration	$\mathbf{+1.00}$ (4/4)	$+0.67$ (1/3)	$+0.50$ (2/3)
Rise time	$\mathbf{+1.00}$ (4/4)	$+0.50$ (4/4)	$\mathbf{+0.88}$ (3/4)
Duration / counts	-0.38 (3/4)	-0.25 (2/4)	$+0.62$ (3/4)
Energy / duration	$+0.38$ (3/4)	$+0.25$ (2/4)	-0.62 (3/4)
Waveform-derived quantities			
Waveform RMS	$\mathbf{+0.88}$ (3/4)	$+0.12$ (1/4)	$\mathbf{-1.00}$ (4/4)
Waveform peak	$\mathbf{+1.00}$ (4/4)	$+0.12$ (1/4)	$\mathbf{-0.84}$ (3/4)
Crest factor	$\mathbf{+0.88}$ (3/4)	-0.38 (1/4)	$+0.50$ (2/4)
Kurtosis	$+0.50$ (2/4)	-0.62 (1/4)	$+0.25$ (2/4)
Peak frequency	-0.41 (3/4)	-0.47 (2/4)	$+0.07$ (4/4)
Spectral centroid	$+0.25$ (2/4)	-0.62 (1/4)	-0.62 (3/4)
Spectral roll-off $95\,\%$	$+0.43$ (4/4)	-0.28 (2/4)	-0.53 (4/4)
Partial power $40\,\mathrm{kHz}$ to $100\,\mathrm{kHz}$	-0.50 (2/4)	$+0.25$ (0/4)	$+0.62$ (1/4)
Partial power $100\,\mathrm{kHz}$ to $300\,\mathrm{kHz}$	$+0.62$ (3/4)	$+0.12$ (1/4)	$\mathbf{-1.00}$ (4/4)
Partial power $300\,\mathrm{kHz}$ to $1000\,\mathrm{kHz}$	$+0.00$ (2/4)	-0.62 (3/4)	-0.25 (0/4)
Retained low-band hits	$1\,886\,543$	$271\,422$	8474
Retained fraction	$100\,\%$	$14.4\,\%$	$0.45\,\%$

Table 4.7: TestBlock 2 low-band speed sensitivity across retained hit populations. Values are mean Spearman correlations across load repeats, except the final two rows.

4.3.6. Assessment of Supported Operational Relationships

Table 4.9 assesses the TB2 speed relationships jointly on association, directional stability and repeatability. The load axis remains withdrawn. Association marks $|\bar{r}_s| \geq 0.7$ in table 4.7. Stability follows from the four load-specific speed correlations: ticks denote the directionally consistent group ($\phi = 1.00$, $s_r = 0.00$–0.25), crosses the less consistent group ($\phi = 0.75$, $s_r = 0.75$–0.95). Repeatability follows table 4.8. Range coverage is common to all features, comprising the three tested stroke periods.

Amplitude, waveform peak, waveform RMS and crest factor satisfy all three criteria. Waveform RMS and crest factor show slightly greater variation between load repeats ($s_r = 0.25$) than amplitude and waveform peak ($s_r = 0.00$), but retain complete directional agreement. Energy and duration are directionally stable and only marginally repeatable. Rise time is perfectly rank ordered, yet differs by about $46\,\%$ between repeated states.

Counts satisfy no criterion. Peak frequency, spectral centroid and spectral roll-off are repeatable but weakly associated and less directionally stable ($\phi = 0.75$, $s_r = 0.80$–0.95). Hit-rate is likewise weakly associated and less directionally stable; its repeatability is not testable, because the ascending and descending passes differ in duration and a rate is not comparable between them.

4.4. Phase 2 TestBlock 3: Contamination Progression

TestBlock 3 uses the same nominal geometry and operating logic as TB2 but introduces progressive grease contamination. Because contamination level is sequentially confounded with elapsed running and possible coupling drift, results are interpreted as condition-associated rather than causal.

4.4.1. Recorded Data Overview

TB3 contains approximately 5.4 million hits across fourteen acquisition files, excluding a separate calibration recording. Activity remains concentrated in the low band and on the same two support-mounted

Quantity	Median diff. [%]	Max. diff. [%]	Verdict
Acquisition quantities			
Amplitude [dB]	0.6	2.5	Repeatable
RMS (acquisition)	2.7	171.8	Repeatable
Energy / duration	6.4	12.6	Repeatable
Duration / counts	9.2	43.2	Repeatable
Duration	13.1	70.0	Marginal
Energy	18.8	74.3	Marginal
Counts	46.2	75.6	Unreliable
Rise time	46.5	109.0	Unreliable
Waveform-derived quantities			
Partial power 40 kHz to 100 kHz	0.0	0.6	Repeatable
Spectral centroid	0.3	2.5	Repeatable
Peak frequency	0.3	3.8	Repeatable
Spectral roll-off 95 %	0.9	13.0	Repeatable
Crest factor	1.7	10.7	Repeatable
Waveform RMS	3.8	10.1	Repeatable
Waveform peak	4.0	14.5	Repeatable
Partial power 300 kHz to 1000 kHz	5.2	31.5	Repeatable
Partial power 100 kHz to 300 kHz	9.3	124.7	Repeatable
Kurtosis	14.4	200.0	Marginal

Table 4.8: TestBlock 2 ascending versus descending repeatability. Values are the median and maximum symmetric relative differences between the two passes across the seven comparable load–speed combinations.

channels as TB2, so sensor-position dependence cannot be assessed. Unlike TB2, however, the mid and high bands retain approximately 145 000 and 61 000 hits, enough to support three-band analysis. Population details are given in section B.3.3.

4.4.2. Experimental Programme and Coverage

The programme comprises an in-block baseline and three increasing Fe contamination levels (table 4.10). Sample 3 contains approximately 6 g of Fe particles in 150 mL of grease; Samples 1 and 2 are successive dilutions. Upper-chamber rollers are replaced between samples while the raceways are retained.

Particles comprise steel and silicon-carbide debris up to approximately 500 µm, coarser than the sub-100 µm particles reported by Miettinen and Andersson [46]. Samples are introduced in increasing concentration without cleaning between stages, so each state builds on the previous one. The contrasting

Feature (low band)	Association	Stability	Repeatability
Amplitude [dB]	✓	✓	✓
Waveform peak	✓	✓	✓
Waveform RMS	✓	✓	✓
Crest factor	✓	✓	✓
Energy	✓	✓	(✓)
Duration	✓	✓	(✓)
Rise time	✓	✓	×
Counts	×	×	×
Peak frequency	×	×	✓
Spectral centroid	×	×	✓
Spectral roll-off 95 %	×	×	✓
Hit-rate	×	×	n.t.

Table 4.9: TestBlock 2 assessment of speed relationships. A parenthesised tick denotes marginal repeatability; n.t. denotes not testable.

Sample	Contamination (volume ratio)	Approx. Fe content	Role
Baseline	Uncontaminated Interflon LS1/2	—	In-block baseline
Sample 1	1:125000	$16\,\mathrm{mg\,kg^{-1}}$	Low contamination
Sample 2	1:2500	$800\,\mathrm{mg\,kg^{-1}}$	Moderate contamination
Sample 3	1:50	$40\,000\,\mathrm{mg\,kg^{-1}}$	High contamination

Table 4.10: TestBlock 3 lubrication contamination levels.

AE response is examined in section 4.4.6.

Acquisition ceilings. Amplitude clipping is negligible ($0.031\,\%$ overall; $0.019\,\%$ in the low band). Duration censoring is more important: $7.98\,\%$ of low-band hits reach the $104\,857.6\,\mu\mathrm{s}$ limit. Duration and duration/counts are therefore treated as bounded quantities; Chapter 6 excludes low-band duration from its primary vector.

The pooled low-band duration censoring fraction does not establish how censoring changes with contamination state. In the four-cycle window table, the fraction of windows whose upper-quartile duration reaches the acquisition ceiling is $64.4\,\%$, $63.0\,\%$, $60.5\,\%$ and $41.7\,\%$ from baseline to Sample 3. This window-level indicator therefore decreases rather than increases at Sample 3 and does not support increasing censoring as a simple explanation for the weakened duration-related load relationships. A raw hit-level censoring-by-run check would be required for a stronger conclusion.

Four contamination runs provide twelve load repeats (run $\times$ speed) and up to sixteen speed repeats (run $\times$ load). The threshold remains stable near $51\,\mathrm{dB}$, so the load axis is interpretable apart from between-recording confounding. Because several files contain multiple load steps, load and speed are reconstructed per hit from operator markers. File mapping and protocol exceptions are given in section B.8.

4.4.3. Influence of Load

Table 4.11 reports pooled load sensitivity across twelve run–speed repeats, for the acquisition quantities and the waveform-derived descriptors. Parentheses give the number of repeats satisfying $|r_s| \geq 0.7$, and bold marks $|\bar{r}_s| \geq 0.7$. The low-band acquisition results are decomposed by contamination state in table 4.13.

Low-band acquisition features are strongly load coupled: median amplitude reaches $+0.97$ with all twelve repeats satisfying the association screen, while energy, counts, duration and both derived ratios have pooled mean correlation magnitudes above 0.8, with at least ten of twelve repeats satisfying $|r_s| \geq 0.7$. Waveform-derived means are much weaker, and neither group shows a comparable pooled relationship in the mid or high bands.

Weak pooled means can, however, mask strong but reversing relationships. Low-band spectral roll-off satisfies the association screen in all twelve repeats yet averages only $+0.14$; crest factor and the $300\,\mathrm{kHz}$ to $1000\,\mathrm{kHz}$ partial power show the same pattern. Stability is therefore determined from the per-run decomposition in section 4.4.5 rather than from pooled means alone.

Sensitivity to retained hit population. A retained-population check was performed for the TB3 acquisition quantities by applying the low-band acquisition-floor threshold before aggregation. This is not an SNR-margin test equivalent to the Phase 4 and TB2 $\mathrm{SNR} \geq 5$ and $\mathrm{SNR} \geq 10$ checks, but tests whether the pooled relationships depend on the lowest-amplitude retained hits. In the low band, median amplitude changes from $+0.97$ to $+0.93$, acquisition RMS from -0.67 to -0.80, energy from $+0.85$ to $+0.93$, counts from $+0.86$ to $+0.95$ and duration from $+0.83$ to $+0.94$. The principal pooled load relationships therefore retain their direction and generally retain or increase their magnitude after the threshold cut.

4.4.4. Influence of Stroke Speed

Table 4.12 reports speed sensitivity across up to sixteen run–load repeats, for the acquisition quantities and the waveform-derived descriptors. Parentheses give the number of repeats satisfying $|r_s| \geq 0.7$,

Quantity	Low band	Mid band	High band
Acquisition quantities			
Hit-rate	-0.63 (7/12)	-0.25 (9/12)	-0.53 (8/12)
RMS (acquisition)	-0.67 (10/12)	-0.20 (8/12)	-0.51 (6/12)
Energy	$\mathbf{+0.85}$ (11/12)	-0.07 (8/12)	$+0.13$ (7/12)
Amplitude [dB]	$\mathbf{+0.97}$ (12/12)	-0.10 (6/12)	-0.14 (5/12)
Counts	$\mathbf{+0.86}$ (11/12)	$+0.05$ (8/12)	$+0.33$ (6/12)
Duration	$\mathbf{+0.83}$ (11/12)	$+0.02$ (8/12)	$+0.38$ (7/12)
Rise time	$+0.25$ (7/12)	$+0.17$ (7/12)	$+0.07$ (2/12)
Duration / counts	$\mathbf{-0.88}$ (10/12)	-0.08 (7/12)	$+0.31$ (6/12)
Energy / duration	$\mathbf{+0.90}$ (10/12)	-0.07 (7/12)	-0.34 (7/12)
Waveform-derived quantities			
Waveform RMS	$+0.00$ (5/12)	$+0.00$ (6/12)	-0.12 (6/12)
Waveform peak	$+0.07$ (9/12)	-0.10 (6/12)	-0.13 (5/12)
Crest factor	-0.10 (11/12)	-0.38 (8/12)	-0.28 (5/12)
Kurtosis	$+0.05$ (7/12)	-0.37 (9/12)	-0.57 (8/12)
Peak frequency	-0.21 (4/11)	-0.20 (10/12)	$+0.34$ (3/12)
Spectral centroid	-0.02 (8/12)	-0.27 (7/12)	-0.24 (4/12)
Spectral roll-off $95\,\%$	$+0.14$ (12/12)	-0.20 (6/12)	-0.36 (3/12)
Partial power $40\,\mathrm{kHz}$ to $100\,\mathrm{kHz}$	$+0.10$ (8/12)	$+0.17$ (5/12)	$+0.63$ (6/12)
Partial power $100\,\mathrm{kHz}$ to $300\,\mathrm{kHz}$	-0.50 (9/12)	-0.17 (5/12)	$+0.23$ (5/12)
Partial power $300\,\mathrm{kHz}$ to $1000\,\mathrm{kHz}$	-0.47 (11/12)	-0.03 (6/12)	-0.23 (5/12)

Table 4.11: TestBlock 3 load sensitivity, pooled across the four contamination runs. Values are mean Spearman correlations with load. Hit-rate is common to both groups and is reported once.

and bold marks $|\bar{r}_s| \geq 0.7$. As in TB2, the three speed states make these coefficients rank-order indicators.

Low-band median amplitude and acquisition RMS move in opposite directions with speed, at $+1.00$ and -1.00 respectively in all sixteen repeats. The two quantities describe different parts of the record: median amplitude is the peak level of the retained transients, whereas $\mathrm{RMS}_{\mathrm{acq}}$ is the background level preceding the threshold crossing. The opposition therefore reports that detected transients become larger while the measured background falls, rather than a single quantity behaving inconsistently. The accompanying low-band changes are an increase in energy, a decrease in duration/counts and an increase in energy/duration, so event morphology also changes with speed. The two RMS quantities are not interchangeable and are reported separately throughout.

4.4.5. Stability Across Contamination States

Table 4.13 decomposes the pooled low-band load correlations by contamination state. Values are means across the three speed repeats, with bold marking $|\bar{r}_s| \geq 0.7$.

Median amplitude remains strongly load coupled ($+0.93$ to $+1.00$) in every state, as do both derived ratios. Energy, counts and duration weaken sharply at Sample 3, while acquisition RMS changes from -0.93 in the baseline to -0.07. Hit-rate coupling strengthens from -0.47 to -0.87. Thus, amplitude and the derived ratios retain stable load relationships, whereas acquisition RMS, energy, counts and duration are condition dependent. A single pooled correction for the latter features would therefore be inappropriate. The available window-level censoring indicator therefore does not support increasing duration censoring as a simple explanation for this weakening (section 4.4.2).

4.4.6. Condition Associated Changes at Comparable Operating States

At matched operating states, the AE population also changes with run order. These differences remain condition-associated because the sequential programme does not isolate contamination from elapsed running.

The four-cycle window construction yields 313 windows from 1356 accepted cycles, corresponding to

Quantity	Low band	Mid band	High band
Acquisition quantities			
Hit-rate	-0.34 (5/16)	$\mathbf{+0.91}$ (15/16)	$\mathbf{+0.77}$ (9/13)
RMS (acquisition)	$\mathbf{-1.00}$ (16/16)	$+0.69$ (10/16)	$+0.63$ (10/15)
Energy	$\mathbf{+0.88}$ (12/16)	-0.34 (9/16)	-0.03 (4/15)
Amplitude [dB]	$\mathbf{+1.00}$ (16/16)	-0.59 (8/16)	-0.27 (5/15)
Counts	$\mathbf{+0.72}$ (11/16)	-0.34 (11/16)	$+0.18$ (4/15)
Duration	$+0.52$ (12/16)	$+0.50$ (10/16)	$+0.20$ (3/15)
Rise time	$+0.22$ (5/16)	$\mathbf{+0.70}$ (11/16)	$+0.20$ (7/15)
Duration / counts	$\mathbf{-1.00}$ (16/16)	$\mathbf{+0.88}$ (12/16)	$+0.40$ (5/15)
Energy / duration	$\mathbf{+1.00}$ (16/16)	$\mathbf{-0.81}$ (10/16)	-0.30 (6/15)
Waveform-derived quantities			
Waveform RMS	-0.22 (7/16)	-0.50 (8/16)	-0.17 (2/15)
Waveform peak	-0.16 (9/16)	-0.62 (8/16)	-0.27 (5/15)
Crest factor	-0.12 (10/16)	-0.66 (7/16)	-0.10 (8/15)
Kurtosis	-0.09 (9/16)	$\mathbf{-0.81}$ (10/16)	-0.03 (8/15)
Peak frequency	$+0.53$ (14/15)	-0.59 (7/16)	$+0.01$ (7/15)
Spectral centroid	$\mathbf{+0.78}$ (15/16)	-0.69 (8/16)	-0.20 (7/15)
Spectral roll-off $95\,\%$	$+0.06$ (15/15)	-0.62 (8/16)	-0.17 (6/15)
Partial power $40\,\mathrm{kHz}$ to $100\,\mathrm{kHz}$	$+0.22$ (13/16)	$\mathbf{+0.72}$ (9/16)	$+0.27$ (9/15)
Partial power $100\,\mathrm{kHz}$ to $300\,\mathrm{kHz}$	$+0.00$ (12/16)	$\mathbf{-0.72}$ (9/16)	$+0.13$ (7/15)
Partial power $300\,\mathrm{kHz}$ to $1000\,\mathrm{kHz}$	$\mathbf{-0.72}$ (11/16)	$+0.41$ (7/16)	-0.13 (7/15)

Table 4.12: TestBlock 3 speed sensitivity, pooled across the four contamination runs. Values are mean Spearman correlations with stroke speed. hit-rate is common to both groups and is reported once.

Quantity (low band)	Baseline	Sample 1	Sample 2	Sample 3
Hit-rate	-0.47	-0.53	-0.67	$\mathbf{-0.87}$
Amplitude [dB]	$\mathbf{+0.93}$	$\mathbf{+0.93}$	$\mathbf{+1.00}$	$\mathbf{+1.00}$
Energy	$\mathbf{+0.93}$	$\mathbf{+0.93}$	$\mathbf{+1.00}$	$+0.53$
RMS (acquisition)	$\mathbf{-0.93}$	$\mathbf{-0.93}$	$\mathbf{-0.73}$	-0.07
Counts	$\mathbf{+0.93}$	$\mathbf{+0.93}$	$\mathbf{+1.00}$	$+0.58$
Duration	$\mathbf{+0.84}$	$\mathbf{+0.91}$	$\mathbf{+0.93}$	$+0.67$
Duration / counts	$\mathbf{-0.87}$	$\mathbf{-0.87}$	$\mathbf{-0.93}$	$\mathbf{-0.87}$
Energy / duration	$\mathbf{+0.87}$	$\mathbf{+0.87}$	$\mathbf{+0.93}$	$\mathbf{+0.93}$

Table 4.13: TestBlock 3 low band load sensitivity per contamination run.

$92.3\,\%$ cycle coverage. Of these, 294 are primary-protocol windows used for the condition-progression comparison and 19 additional Sample 1 windows are retained as repeats. Across the accepted window spans, $79.9\,\%$ of AE hits are assigned.

Hit activity increases consistently with run order. Between baseline and Sample 3, time-based hit-rate rises from 6973 to $13\,241$ hits per minute at $100\,\mathrm{kN}$, from 5562 to $10\,577$ at $260\,\mathrm{kN}$, from 2721 to 5103 at $785\,\mathrm{kN}$, and from 1916 to 3854 at $1305\,\mathrm{kN}$. Mean duration falls from $13\,407\,\mu\mathrm{s}$ to $6921\,\mu\mathrm{s}$ and counts from 491 to 233; amplitude changes little ($57.5\,\mathrm{dB}\rightarrow58.9\,\mathrm{dB}$), while RMS decreases ($125.7\,\mu\mathrm{V}\rightarrow100.9\,\mu\mathrm{V}$). The later state therefore contains more, shorter events at similar amplitude.

The run-order ranking is robust to the selected window size. The low-band trend gives $\rho = 0.8$ for one-cycle windows and $\rho = 1.0$ for three-, four- and five-cycle windows. The high-band trend gives $\rho \approx 0.95$ for one- and three-cycle windows and $\rho = 1.0$ for four- and five-cycle windows.

Frequency-band activity also broadens. High-band activity rises from a median of 0–0.5 hits per window in baseline/Sample 1 to 250 in Sample 3; the mid band rises from 54 to 641. Low-band activity increases by roughly a factor of 2.5. The three-band population in TB3 is therefore itself condition associated.

This trend differs from the contamination response reported by Miettinen and Andersson [46]: TB3 shows higher hit-rate with little amplitude change. The direction is consistent across runs and loads and is not explained by threshold drift. The coarser particles used here may limit transfer of the published relationship, but the mechanism is not established.

4.4.7. Assessment of Supported Operational Relationships

Table 4.14 focuses on the criterion for which TB3 provides distinct evidence: stability of the operational relationships across contamination states. Association is reported in tables 4.11 and 4.12. Load stability is taken from the state-specific decomposition in table 4.13, while speed stability is judged across contamination states.

TB3 contains revisited commanded states in several operating blocks, so repeatability is not intrinsically untestable. A Chapter 4 feature-level ascending–descending repeatability assessment was not performed for TB3; the present assessment therefore reports stability across contamination states rather than a repeatability verdict. Relationships that collapse, weaken or change systematically with contamination are treated as condition dependent rather than pooled correction targets.

Feature (band)	Axis	Stable across contamination
Amplitude (low)	Load	✓
Duration / counts (low)	Load	✓
Energy / duration (low)	Load	✓
Energy (low)	Load	×
Counts (low)	Load	×
Duration (low)	Load	×
RMS (acquisition, low)	Load	×
Hit-rate (low)	Load	(×)
Hit-rate (mid, high)	Speed	✓
Spectral centroid (low)	Speed	✓
Kurtosis (mid)	Speed	✓

Table 4.14: TestBlock 3 stability of candidate operational relationships across contamination states. Ticks mark relationships that remain stable; crosses mark condition-dependent behaviour. The parenthesised cross marks a relationship that strengthens with contamination rather than collapsing.

Supported TB3 load relationships are low-band amplitude and the two derived ratios. Energy and counts lose their load relationship at Sample 3, duration weakens, and acquisition RMS collapses. Low-band hit-rate instead strengthens with run order, indicating condition-dependent coupling rather than a stable load relationship. Stable speed relationships are mid- and high-band hit-rate, low-band spectral centroid and mid-band kurtosis. Contamination therefore remains an interpretation axis rather than a correction variable.

4.5. Cross Dataset Synthesis

4.5.1. Distribution of Operational Sensitivity Across the Feature Set

Operational sensitivity is feature-, band- and dataset-specific. Median amplitude is the most consistent descriptor: it is load associated in Phase 4, speed ordered in TB2, and load associated in all TB3 repeats. Waveform RMS and peak are strong in Phase 4 but show weak pooled TB3 load relationships with substantial reversal between repeats, so their relationships do not transfer directly.

Acquisition RMS and waveform RMS must remain separate. They carry opposite signs in the Phase 4 low and mid bands and describe different parts of the record: the background preceding detection, and the stored transient itself. hit-rate is similarly context dependent: it is strongly load associated in the Phase 4 high band, weakly speed associated in unfiltered TB2, strongly speed associated in TB3 mid/high bands, and increasingly negative with TB3 load as contamination progresses.

Single-value spectral descriptors often have weak pooled means because their directions reverse between repeats. Low-band TB3 roll-off, for example, clears the association screen repeatedly but averages near zero. The key distinction is therefore between association, repeatability and directional

stability: a feature can satisfy the first two and still be unsuitable for correction if its operational relationship changes state.

4.5.2. Stability, Repeatability and Coverage

The datasets provide complementary evidence. Phase 4 tests load association and cross-test stability; TB2 adds within-session repeatability; TB3 tests stability across condition states. Together they show that association is necessary but insufficient: supported operational relationships must also persist across the variation available in the experiment.

Figure 4.4 illustrates this distinction. Each relationship is standardised within its own operating points, so the panels compare direction and form rather than absolute magnitude. Panel (a) shows the stable Phase 4 mid-band waveform-RMS load response; panel (b) shows TB2 amplitude ordered consistently with speed. Panel (c) shows TB3 amplitude retaining its positive load response across contamination states. Panel (d) contrasts this with acquisition RMS, whose state-mean load correlation weakens from -0.93 to -0.07. Its pooled mean therefore hides a state-dependent relationship, demonstrating why stability is assessed separately from association.

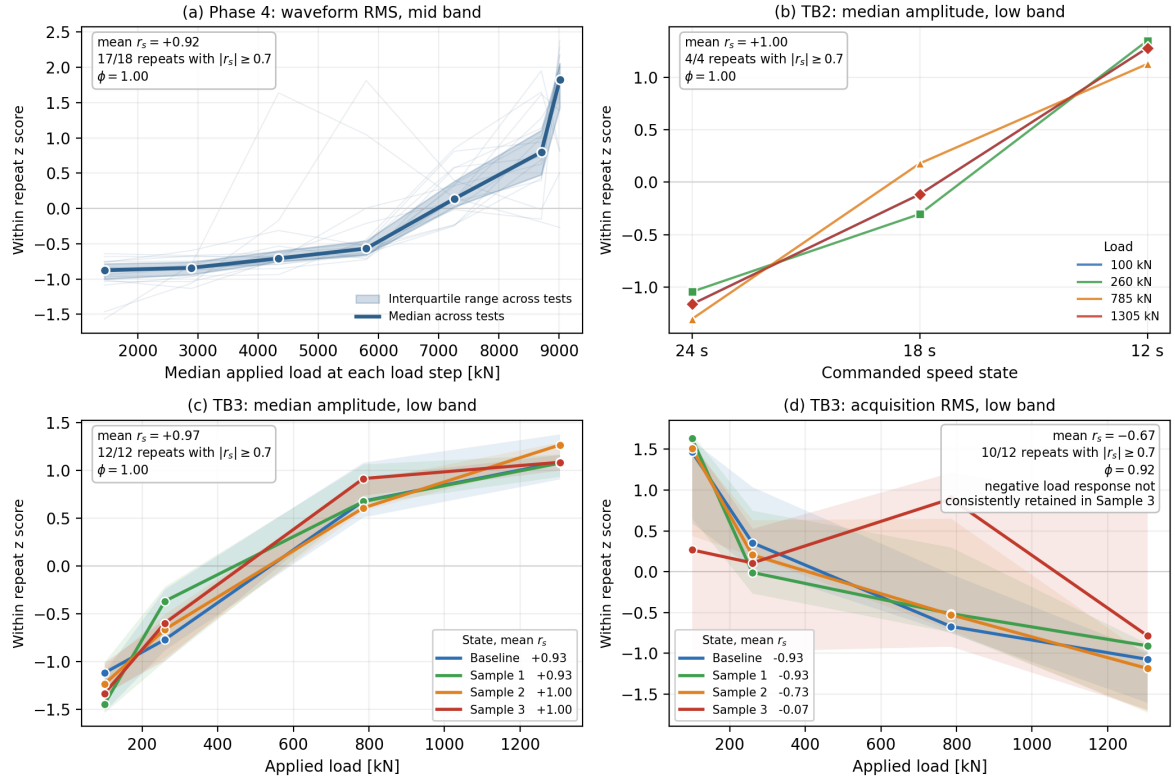


Figure 4.4: Representative operational relationships illustrating the Chapter 4 stability assessment.

Shaded regions show between-repeat variation, not confidence intervals; TB3 ranges use the full minimum–maximum span because each state contains only three repeats. Correlation statistics are calculated from the underlying individual relationships. Standardisation is used only for visual comparison and is not the correction applied in Chapter 5.

Two limitations are common across datasets. First, acquisition ceilings bound amplitude- and duration-related quantities: duration censoring affects $5.29\,\%$ of Phase 4 hits, $1.40\,\%$ of TB2 low-band hits and $7.98\,\%$ of TB3 low-band hits. Second, no dataset provides both clean within-recording load and speed axes. Cross-dataset agreement therefore strengthens evidence, but numerical relationships are not assumed to transfer between rigs.

4.5.3. Implications for Operational Correction

The correction stage carries forward three principles. Stable amplitude-related quantities and selected derived ratios provide prior candidates, but each dataset is reassessed independently. Relationships that depend on unmeasured state, as seen for TB3 acquisition RMS, energy and counts, are poor correction targets. Chapter 5 therefore performs the full correction and validation independently, while Chapter 6 examines the resulting multivariate structure. Table 4.15 summarises the evidence supporting that transition.

4.6. Chapter Summary

This chapter shows that operational sensitivity is strongly feature- and band-dependent, and that correlation strength alone is insufficient for correction eligibility.

Phase 4 provides the strongest load evidence. High-band hit-rate, amplitude anchors and several mid-band energy-, count- and duration-related features are load associated, while speed, lubrication and SNR sensitivity qualify their interpretation. TB2 contributes a clean speed axis and the chapter's direct repeatability test; its load axis is withdrawn because threshold and load are confounded. TB3 provides the strongest stability test across changing condition: low-band amplitude and the derived ratios retain their load relationships, whereas acquisition RMS, energy, counts and duration weaken at high contamination. At matched operating states, later contamination runs also show more, shorter events and broader mid- and high-band activity.

Across the datasets, supported operational relationships are those that remain associated and stable, and repeatable where the design permits. The amplitude anchors and selected derived ratios provide the strongest prior evidence of stable operational dependence, but no feature is transferred directly into correction. Hit-rate carries operational information whose meaning depends on axis, band and retained population. Single-value spectral descriptors are generally weaker or less stable than the amplitude-related quantities, although individual supported relationships remain. Because none of the linear bearing datasets combines a suitable reference condition, sufficiently clean multivariable operational coverage and an independent validation structure, the complete predictive correction procedure is not applied here. Chapter 5 therefore reassesses the full-scale feature set independently, while Chapter 6 examines the remaining multivariate structure.

Dataset and axis	Evidence quality	Main observed relationship	Stability or repeatability	Implication for operational modelling
Phase 4 load	Strong	Several feature band combinations load associated, strongest for amplitude anchors and mid band energy	Repeatability not testable; stable across sweeps, with selected relationships SNR robust	Supported relationships in an advanced degradation state
Phase 4 speed	Moderate	Feature levels change with speed; event rate and amplitude decouple in the high band	Between recording comparison	Descriptive modifier
Phase 4 lubrication	Moderate	Feature and band dependent influence	Not directly measured	Descriptive modifier
TB2 load	Invalid	Threshold and load confounded	No interpretation	Withdrawn; no modelling
TB2 speed	Strong within recording, narrow range	Several features consistently ordered with speed	Repeatability available	Supported where association, stability and repeatability align
TB3 load, scalar	Strong	Amplitude load coupled at $+0.97$ in all twelve repeats; energy, counts, duration and the derived ratios have correlation magnitudes above 0.8 in the low band	Amplitude and the two ratios stable across all four contamination states; energy, acquisition RMS, counts and duration lose coupling at the highest level	Amplitude and the derived ratios are the defensible targets; the remainder are condition dependent
TB3 load, waveform	Weak pooled	Coupled in most individual repeats but direction reverses between them, so pooled means cancel to near zero	Directional instability, not absence of response	Do not fit one pooled relationship where direction changes
TB3 speed	Moderate	Some feature band combinations stable across contamination	Assessed across contamination states	Supported where associated and stable across contamination
TB3 contamination	Condition progression	AE population changes with contamination	Sequential design limits causal reading	Condition associated evidence for later interpretation, not correction

Table 4.15: Cross dataset synthesis of operational evidence.

5

Full-Scale Swivel Bearing: Experiment and Operational Influence

Chapter 4 characterised operational effects in controlled linear-bearing rigs, but their transfer to a full-scale rotary bearing remained uncertain. This chapter examines the full-scale swivel-bearing experiment at the Bluewater Test Facility (BTF), which provides the reference condition, a bearing without known damage, for the correction framework of Chapter 3. It quantifies the influence of applied pressure and rotation speed on AE features and, because the BTF design permits held-out validation, applies the complete correction-eligibility sequence.

The test matrix separates the two operational factors: pressure is fixed within each recording and stepped between recordings, while each recording covers all speed magnitudes in both directions. Of ten recordings collected over two days, the analysis uses the six-recording same-day sweep from 25 to $0\ \mathrm{barg}$. Because the response changes form above $15\ \mathrm{barg}$, the sensitivity tables report the reduced 0–$15\ \mathrm{barg}$ sweep, which is also the range of the correction models; complete-sweep estimates are given in section C.8 and quoted where they differ.

5.1. Test Bearing and Setup

The test object was a full-scale buoy swivel bearing used in CALM buoy and turret mooring systems. It contains three roller rows (figure 5.1) carrying axial load, uplift and moment reversal, and centring the assembly. Load is distributed across multiple contacts, so AE represents combined source activity. Without a detailed load-distribution model, the analysis uses rank-order relationships with applied hydraulic pressure. The outer-ring sensor position also introduces a substantial steel transmission path that may attenuate higher-frequency content. Dimensions are given in Appendix C.2.

No defect was introduced or known to be present. The bearing is therefore treated as a mature, run-in reference, with approximately 2.5×10^5 equivalent revolutions and Castrol Spheerol EPL 2 grease. The interior was not opened for inspection. An axial-clearance check was within the allowable range, although the underlying data were unavailable. The dataset therefore represents a bearing without known damage, not proof of an unworn bearing.

The swivel was mounted with its rotation axis vertical. The lower half rotated on a turntable while the upper half remained stationary. A manually initiated electric motor applied five commanded speed levels at each pressure. Testing was indoors at approximately $23\ ^\circ\mathrm{C}$.

Axial load was applied hydrostatically. The internal volume of the swivel is sealed between the stationary upper half and the rotating lower half and closed by flanges. This volume was filled with water through two inlets on opposite sides and then pressurised. Once the target pressure was reached, the water was static, so the pressure was the same throughout the volume and the position of the inlets did not affect the load.

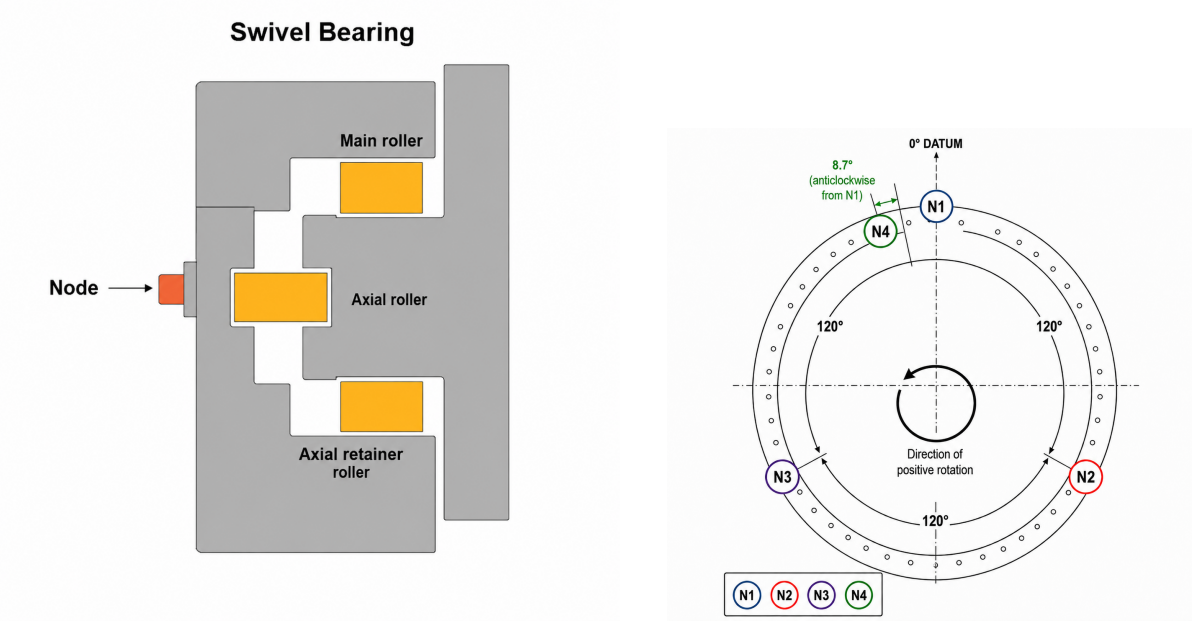


Figure 5.1: BTF swivel bearing: cross-section (left) and AE sensor-node layout (right).

The pressure pushes perpendicular to every internal surface. Because the cavity is axisymmetric, the radial components cancel, while the axial components push the two halves apart. The resulting separating force equals the gauge pressure multiplied by the area enclosed by the outermost seal (equation (5.1)). The two halves are held together only by the bearing, so this force is carried by the roller rows that resist axial separation, not by the turntable or the support frame. Because the pressure acts equally around the circumference, the rows are loaded uniformly and no rotating load zone is formed. The load could therefore be changed simply by adjusting the pressure, without modifying the assembly or the sensor mounting. Pressure is reported in bar gauge (barg); $0\ \mathrm{barg}$ is ambient pressure. The hydraulic separating force is

$$F_z = p_{\mathrm{g}} A_{\mathrm{eff}}, \qquad A_{\mathrm{eff}} = \frac{\pi D_{\mathrm{seal}}^2}{4} = 726842\ \mathrm{mm}^2, \tag{5.1}$$

with $D_{\mathrm{seal}} = 962\ \mathrm{mm}$, giving

$$F_z = 72.68\, p_{\mathrm{g}}, \tag{5.2}$$

where F_z is in kN and p_{g} in barg; the resulting forces are listed in table 5.1. The rotating assembly also contributed approximately $2275\ \mathrm{kg}$ of dead weight, so $0\ \mathrm{barg}$ is not zero total load. This constant offset does not change rank order. Because force is linear in pressure, Spearman correlation is identical for either variable; gauge pressure is therefore retained.

The approved range was 0–$25\ \mathrm{barg}$. Minor flange weeping was reported at 20 and $25\ \mathrm{barg}$, where pressure was manually adjusted; no leakage was reported at $15\ \mathrm{barg}$ or below. The available pressure traces are not fully calibrated and are used only to confirm the sequence (Appendix C.3).

Four AE nodes were installed on the rotating outer ring. N1–N3, at $0°$, $120°$ and $240°$, provide channels 1–9 used in this analysis; N4 is excluded because it used a different sensor system. Positive rotation is anticlockwise. Each retained node carried an R6α, R15α and WSα transducer with the same nominal responses and low-, mid- and high-band passbands as Phase 4 (section 4.2.3), amplified by a Vallen AEP5H at $40\ \mathrm{dB}$. Magnetic mounting and silicone-gel coupling were unchanged throughout the selected sequence. Pencil-lead break (Hsu–Nielsen) sources were applied after mounting to verify coupling and channel sensitivity before acquisition began.

Acquisition followed the HiTeAM Phase 4 configuration. Transients were sampled at $5\ \mathrm{MHz}$ with a $200\ \mu\mathrm{s}$ pre-trigger segment and variable length up to $1638\ \mu\mathrm{s}$, stored as FLAC-compressed 16-bit integers. Hardware triggering used $45\ \mathrm{dB}$; amplitudes are reported relative to $1\ \mu\mathrm{V}$. The higher band-specific

thresholds of section 5.3.1 are post-hoc screening criteria. Each recording contains a `.pridb` feature database and `.tradb` waveform database.

5.2. Test Programme and Data Preparation

The campaign comprised ten recordings on 7–8 July 2026 (Appendix C.1). First-day recordings had less stable pressure and different background acoustic conditions, so they are retained for data-quality context but excluded from the operational analysis. Each pressure condition used five nominal speeds, $\{0.5, 1.5, 3.0, 4.5, 6.0\}$ $\mathrm{deg/s}$, each run first anticlockwise and then clockwise. This gave ten passes per recording and repeated visits to each commanded state (section 5.5). Speed is retained in $\mathrm{deg/s}$, the commanded quantity, because conversion to tangential velocity would require an arbitrary reference radius and would not alter the rank correlations.

The six recordings from 8 July 2026 form a complete 0–25 barg sweep with all nine channels active (table 5.1). Pressure decreased from 25 to 0 barg. Hit yield ranges from 609 to $77\,311$, while median amplitude remains within 51.6–53.4 dB.

Test case	Pressure (barg)	F_z (kN)	Duration (min)	Total hits	Median amplitude (dB)
SB-L7-SR	0	0	44.4	3332	53.2
SB-L6-SR	5	363	43.2	15797	53.3
SB-L5-SR	10	727	42.8	77311	53.4
SB-L4-SR	15	1090	45.9	76388	52.8
SB-L3-SR	20	1454	44.1	11576	51.6
SB-L2-SR	25	1817	45.0	609	52.8

Table 5.1: BTF recordings used in the analysis.

The BTF analysis follows the shared pipeline of Chapter 3; its stage-by-stage implementation is given in table C.5. Unlike Chapter 4, the BTF experiment had no continuous operational data stream: pressure is assigned per recording and speed from the minute-resolution operator log. The analysis unit is the pressure–speed–channel combination, 270 in total, with activity expressed both per minute and per nominal rotation.

5.2.1. Speed-Step Time Alignment

Passes were started manually, producing variable delays. Boundaries inferred from commanded speed and angle drifted, while AE quiet intervals did not reliably identify reversals. The operator rotation log was therefore used to assign pass start times and signed speeds (Appendix C.4). For test j, the logged start time of pass k and the timestamp of hit i are expressed relative to the first AE hit $T_{0,j}^{\mathrm{AE}}$:

$$\tau_{j,k} = \frac{T_{j,k}^{\log} - T_{0,j}^{\mathrm{AE}}}{60\ \mathrm{s}}, \qquad \tau_{i,j} = \frac{T_{i,j}^{\mathrm{AE}} - T_{0,j}^{\mathrm{AE}}}{60\ \mathrm{s}}. \tag{5.3}$$

Only pass starts were logged, so pass k is assigned $[\tau_{j,k},\, \tau_{j,k+1})$, with the final window ending at the last hit. This gives ten contiguous windows per test. Anticlockwise and clockwise windows are pooled at each nominal speed, and hit-rate is count divided by elapsed window duration. The method avoids assumptions about rotation stop times but retains the stationary interval after each pass (section 5.7).

5.2.2. Waveform Decoding and Feature Extraction

Waveforms are decoded using the stored sampling frequency and sample count. All were sampled at $5\ \mathrm{MHz}$, with 1616–8192 samples and a 1000-sample pre-trigger in almost all cases. Invalid records were defined as fewer than 16 samples or all zero; none occurred, and all $185\,013$ waveforms decoded successfully. Decoding precedes amplitude screening, so hit and waveform populations remain aligned.

Per waveform, RMS, peak amplitude, crest factor and bias-corrected excess kurtosis are computed in the time domain. RMS and peak remain in sample units because a common linear scale does not affect Spearman correlation. The unwindowed FFT provides peak frequency, spectral centroid, $95\,\%$ roll-off and the LF, MF and HF partial powers in the three instrument passbands. These passbands overlap

and do not span the full spectrum, so the partial powers are not complementary and are computed for every waveform irrespective of transducer.

The unwindowed FFT provides peak frequency, spectral centroid, $95\,\%$ roll-off and the LF, MF and HF partial powers over the three instrument passbands ($40\,\mathrm{kHz}$ to $100\,\mathrm{kHz}$, $95\,\mathrm{kHz}$ to $180\,\mathrm{kHz}$ and $180\,\mathrm{kHz}$ to $580\,\mathrm{kHz}$).

5.3. Recorded Population and Amplitude Screening

The six recordings contain $185\,013$ hits: $136\,219$ mid-band, $39\,820$ low-band and 8974 high-band. N2 contributes $63.1\,\%$, N3 $23.0\,\%$ and N1 $13.9\,\%$. Channel 5 alone supplies $55.5\,\%$ and channel 7 a further $16.6\,\%$; almost all high-band hits come from channel 6. These values describe the unscreened population. Figure 5.2 reports the within-recording share by frequency band (left) and sensor node (right).

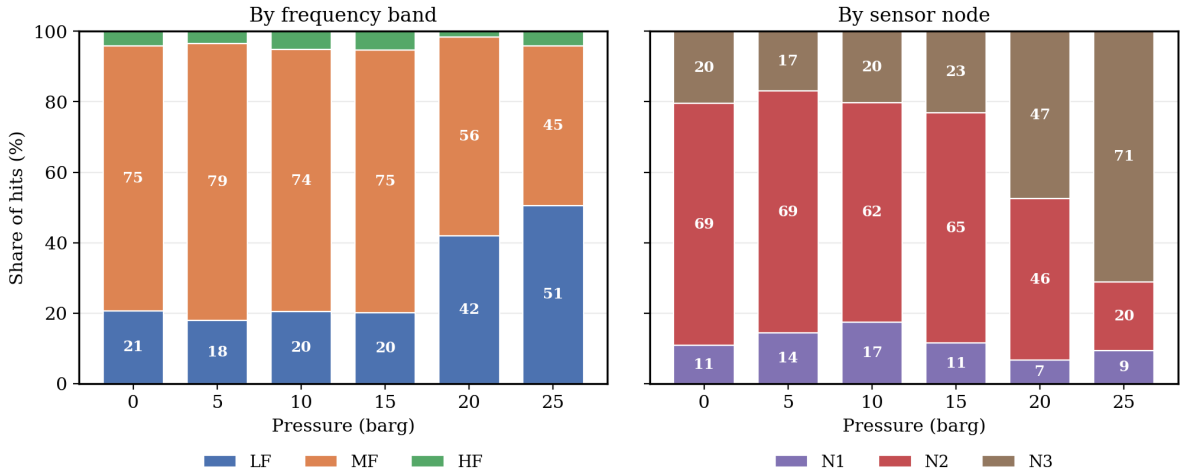


Figure 5.2: Hit composition by pressure.

Composition is stable from 0 to 15 barg but changes at higher pressure. The mid band contributes $74\,\%$ to $79\,\%$ and N2 $62\,\%$ to $69\,\%$ over the lower four recordings. At 20 barg, the low band and N3 rise to $41.9\,\%$ and $47.4\,\%$; at 25 barg they reach $50.6\,\%$ and $71.1\,\%$. The high-pressure change therefore concerns both activity and its sensor and band distribution.

Activity also varies strongly through each recording (Appendix C.6). The 10 and 15 barg tests contain $83\,\%$ of all hits. At 0–15 barg, activity occurs at every speed and band ordering is preserved; at 20 and 25 barg, it shifts strongly towards the two fastest levels, which occupy only $15\,\%$ of elapsed time. Hit-rate spans 13.5–$1806\,\mathrm{hits/min}$, whereas median amplitude remains within 51.6–$53.4\,\mathrm{dB}$ and band medians change by only 1.3–$2.9\,\mathrm{dB}$ across the full pressure range. Pressure therefore changes detected activity far more than event amplitude.

5.3.1. Amplitude Screening

Hardware triggering at $45\,\mathrm{dB}$ is consistent with the quietest recorded hit at $45.9\,\mathrm{dB}$. Neither acquisition ceiling is reached: the maximum amplitude remains below $49\,990\,\mu\mathrm{V}$, and the longest duration is $22\,475\,\mu\mathrm{s}$, well below the $104\,857.6\,\mu\mathrm{s}$ limit. The `.pridb` duration is a threshold-crossing quantity and is distinct from waveform record length.

A second flag, `snr_pass`, marks whether a hit exceeds a band-specific amplitude threshold,

$$\texttt{snr_pass}_i = \begin{cases} \text{True}, & A_{\mathrm{dB},i} \geq A_{\mathrm{th},b}, \\ \text{False}, & \text{otherwise}, \end{cases} \tag{5.4}$$

for hit i in band b, with $A_{\mathrm{th},b} = 55\,\mathrm{dB}$ in the low band and $50\,\mathrm{dB}$ in the mid and high bands, that is 10 and $5\,\mathrm{dB}$ above the trigger. Despite the flag name, this is an absolute amplitude gate; no per-hit noise level is measured.

The screen is applied throughout the chapter. It removes the $20\ \mathrm{barg}$ composition anomaly associated with near-threshold hits: the low-band share falls from $41.9\,\%$ to $12.2\,\%$ and N3 from $47.4\,\%$ to $21.8\,\%$, approaching the lower-pressure composition. The cost is uneven. The screen removes $69.3\,\%$ of low-band, $15.6\,\%$ of mid-band and $7.7\,\%$ of high-band hits, retaining $135\,481$ of $185\,013$ ($73.2\,\%$). The retained population becomes $84.8\,\%$ mid-band, with channel 5 alone supplying $65.8\,\%$. This follows from an amplitude distribution concentrated close to the $45\ \mathrm{dB}$ trigger.

Sensitivity to the screen was tested with no screen, the adopted thresholds, and thresholds raised uniformly by 5 and $10\ \mathrm{dB}$ (tables C.6 and C.7). Mid-band hit-rate against pressure is essentially unchanged across all four settings. Mid-band crest factor and kurtosis against pressure are strongest at the adopted threshold (-0.77 and -0.81) and weaken on either side, so their apparent coupling is threshold-dependent. Low-band waveform peak against speed weakens progressively as screening removes the low-band population, from -0.92 unscreened to -0.20 at $+10\ \mathrm{dB}$, whereas mid-band waveform RMS retains its sign and approximate magnitude. The adopted threshold is therefore a compromise: removing it restores the $20\ \mathrm{barg}$ near-threshold anomaly, while raising it leaves only $8.2\,\%$ of low-band hits at $+5\ \mathrm{dB}$ and $2.0\,\%$ at $+10\ \mathrm{dB}$. It is not treated as an optimal setting, and threshold-sensitive relationships remain qualified.

5.4. Operational Sensitivity

All results use the $135\,481$ retained hits. Pressure sensitivity is estimated by correlating the pressure-specific medians at each of the five nominal speeds, giving five Spearman coefficients per feature and band. Speed sensitivity is estimated from the five speed-specific medians within each recording, giving one coefficient per recording. In the reduced sweep, each pressure coefficient therefore rests on four recordings and each speed mean on four coefficients. Anticlockwise and clockwise passes are pooled at each speed; section 5.5 shows that this is supported for per-hit descriptors but not for hit-rate.

Tables 5.2 and 5.3 report the signed mean correlation; parentheses give the number of individual coefficients with $|r_s| \geq 0.7$, and bold marks $|\bar{r}_s| \geq 0.7$. The estimates are not fully independent and are interpreted as descriptive trends. Median hit amplitude and median waveform peak describe the same transient maximum, and Spearman correlation is unaffected by the monotonic conversion to decibels, so their coefficients coincide or nearly coincide. Lubrication condition was not varied or measured at the BTF; any lubrication-mediated effect therefore remains within the unexplained pressure- and speed-related variability.

5.4.1. Pressure

Pressure is linearly proportional to F_z. Over the complete six-recording sweep, no hit-feature mean reaches $|\bar{r}_s| \geq 0.7$; the largest are mid-band rise time ($+0.68$), duration ($+0.47$) and counts ($+0.46$). The strongest waveform relationships are mid-band crest factor (-0.77) and kurtosis (-0.81), each coupled at four of five speed levels, followed by high-band spectral roll-off ($+0.68$); waveform RMS and peak remain weak in every band (tables C.8 and C.9).

Figure 5.3 shows why the complete sweep is weak for activity: hit-rate is unimodal across pressure, rising to 10–$15\ \mathrm{barg}$ and then falling. Filled markers identify the reduced 0–$15\ \mathrm{barg}$ sweep; open markers show the excluded 20 and $25\ \mathrm{barg}$ recordings. A symmetric-logarithmic vertical scale allows all bands and the zero-hit point to share one axis.

Within the reduced sweep (table 5.2), hit-rate rises to $+0.88$–$+0.96$ and is coupled at all five speeds in every band. Mid-band duration, energy, rise time and counts reach $+0.80$–$+0.86$, while amplitude remains near zero. Among the waveform features, mid-band RMS increases to $+0.60$, high-band RMS to $+0.44$ and the HF partial power of mid-band transients to $+0.72$, whereas mid-band crest factor and kurtosis weaken to -0.32 and -0.48. Some large coupling counts coexist with small signed means because individual correlations reverse sign. The exclusion therefore affects feature groups differently rather than uniformly strengthening all relationships. The hit-rate trend is consistent in direction but must be read with its poor directional repeatability (section 5.5).

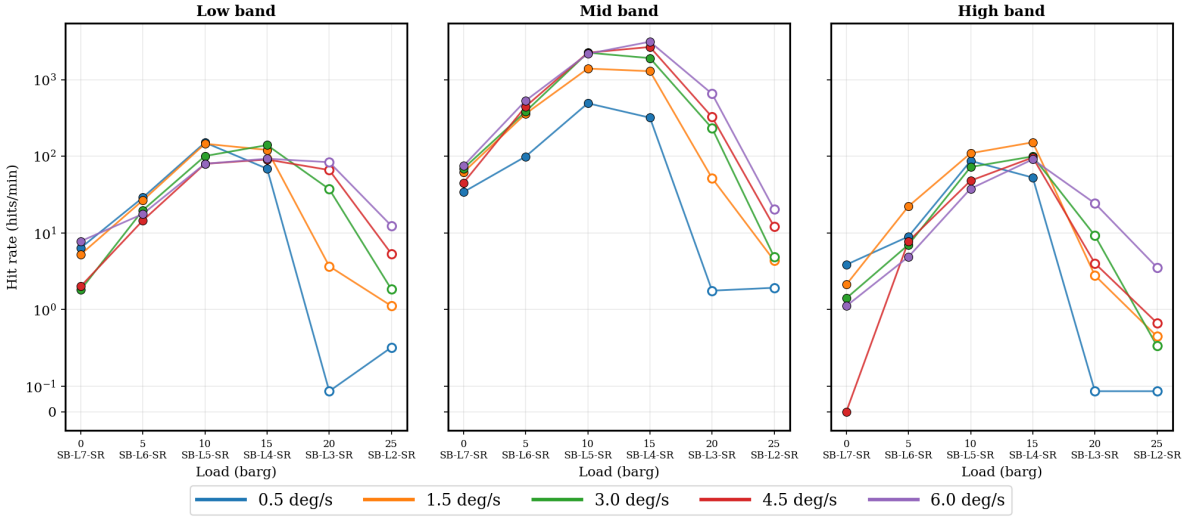


Figure 5.3: Hit-rate versus pressure. Filled markers indicate the measurements retained for the reduced 0–15 barg sweep used in the main analysis. Open markers indicate the excluded 20 and 25 barg recordings, during which hydraulic leakage occurred.

Quantity	Low band	Mid band	High band
Acquisition quantities			
Hit-rate	$+\mathbf{0.92}$ (5/5)	$+\mathbf{0.88}$ (5/5)	$+\mathbf{0.96}$ (5/5)
Energy	-0.08 (1/5)	$+\mathbf{0.80}$ (4/5)	$+0.30$ (2/5)
Amplitude (dB)	-0.20 (1/5)	$+0.04$ (3/5)	$+0.06$ (0/5)
Counts	$+0.13$ (1/5)	$+\mathbf{0.86}$ (5/5)	$+0.22$ (3/5)
Duration	$+0.28$ (1/5)	$+\mathbf{0.80}$ (4/5)	$+0.12$ (3/5)
Rise time	$+0.20$ (1/5)	$+\mathbf{0.80}$ (4/5)	-0.50 (3/5)
Duration / counts	$+0.24$ (1/5)	$+0.56$ (5/5)	-0.56 (2/5)
Energy / duration	-0.20 (2/5)	-0.20 (5/5)	$+0.16$ (3/5)
Waveform-derived quantities			
Waveform RMS	-0.28 (1/5)	$+0.60$ (3/5)	$+0.44$ (2/5)
Waveform peak	-0.20 (1/5)	$+0.04$ (3/5)	$+0.06$ (0/5)
Crest factor	$+0.36$ (1/5)	-0.32 (3/5)	-0.00 (3/5)
Kurtosis	$+0.24$ (0/5)	-0.48 (4/5)	-0.06 (3/5)
Peak frequency	-0.50 (3/5)	$+0.40$ (1/5)	$+0.22$ (3/5)
Spectral centroid	-0.04 (2/5)	$+0.64$ (2/5)	$+0.22$ (3/5)
Spectral roll-off	$+0.56$ (2/5)	$+0.52$ (1/5)	$+0.40$ (4/5)
LF partial power	-0.52 (3/5)	-0.12 (2/5)	-0.38 (3/5)
MF partial power	$+0.68$ (3/5)	-0.16 (1/5)	$+0.10$ (0/5)
HF partial power	$+0.04$ (1/5)	$+\mathbf{0.72}$ (3/5)	$+0.06$ (1/5)

Table 5.2: BTF pressure sensitivity, 0–15 barg.

5.4.2. Rotation Speed and Exposure

Each speed level contains two full rotations but occupies less time as speed increases, from about $26\ \mathrm{min}$ at $0.5\ \mathrm{deg/s}$ to $3.9\ \mathrm{min}$ at $6.0\ \mathrm{deg/s}$. Hit-rate is therefore reported per minute and per nominal rotation. In the mid band, the median time-based rate rises from about 19 to $56\ \mathrm{hits/min}$ between 0.5 and $6.0\ \mathrm{deg/s}$, whereas activity per rotation falls from about 229 to $108\ \mathrm{hits/rotation}$. Hits per rotation therefore reduces the exposure-time effect but is not speed-independent. It remains a nominal measure because angular displacement is inferred from the operator log.

Over the complete sweep, mid-band hit-rate per minute has the strongest speed relationship ($+0.88$), followed by low-band duration/counts ($+0.80$). Hit-rate per rotation is negative in all three bands (-0.25, -0.18 and -0.43), showing that the positive per-minute trend largely reflects shorter exposure time rather than greater emission per unit angular travel. Amplitude, waveform RMS and waveform peak are negative in all bands, and the partial powers change with speed (tables C.10 and C.11).

Within the reduced sweep (table 5.3), most relationships strengthen. Mid-band energy and counts reach -0.95 and -0.94, amplitude is strongly negative in all bands, and waveform RMS and peak strengthen accordingly. Mid-band hit-rate per minute remains positive at $+0.82$, while per-rotation hit-rate is strongly negative in the low and high bands. Faster motion therefore produces fewer, less energetic events per nominal rotation even though mid-band events occur more rapidly in time. Within low-band transients, MF partial power falls to -0.90, HF partial power rises to $+0.80$, and centroid and roll-off decrease, indicating a redistribution of spectral content rather than a uniform amplitude effect. With only four recordings contributing to each mean, this spectral pattern remains provisional.

Quantity	Low band	Mid band	High band
Acquisition quantities			
Hit-rate (per minute)	-0.38 (2/4)	$\mathbf{+0.82}$ (3/4)	-0.65 (3/4)
Hit-rate (per rotation)	$\mathbf{-0.87}$ (4/4)	-0.52 (2/4)	$\mathbf{-0.92}$ (4/4)
Energy	-0.50 (2/4)	$\mathbf{-0.95}$ (4/4)	$\mathbf{-0.70}$ (3/4)
Amplitude (dB)	$\mathbf{-0.93}$ (4/4)	$\mathbf{-0.82}$ (3/4)	$\mathbf{-0.75}$ (3/4)
Counts	-0.37 (2/4)	$\mathbf{-0.94}$ (4/4)	-0.42 (2/4)
Duration	-0.20 (1/4)	-0.62 (2/4)	-0.38 (3/4)
Rise time	$+0.07$ (2/4)	-0.53 (3/4)	-0.03 (2/4)
Duration / counts	$\mathbf{+0.93}$ (4/4)	$\mathbf{+0.85}$ (4/4)	$+0.47$ (2/4)
Energy / duration	$\mathbf{-0.78}$ (2/4)	$\mathbf{-0.93}$ (4/4)	-0.12 (3/4)
Waveform-derived quantities			
Waveform RMS	-0.60 (3/4)	$\mathbf{-0.95}$ (4/4)	$\mathbf{-0.82}$ (3/4)
Waveform peak	$\mathbf{-0.93}$ (4/4)	$\mathbf{-0.87}$ (3/4)	$\mathbf{-0.72}$ (3/4)
Crest factor	-0.12 (0/4)	-0.47 (1/4)	-0.15 (1/4)
Kurtosis	-0.23 (1/4)	-0.60 (2/4)	$+0.05$ (0/4)
Peak frequency	$+0.01$ (0/4)	$+0.68$ (3/4)	-0.40 (2/4)
Spectral centroid	$\mathbf{-0.72}$ (3/4)	-0.38 (1/4)	-0.03 (2/4)
Spectral roll-off	$\mathbf{-0.80}$ (3/4)	$+0.24$ (0/4)	$+0.05$ (2/4)
LF partial power	$+0.40$ (1/4)	$\mathbf{+0.75}$ (3/4)	$+0.20$ (2/4)
MF partial power	$\mathbf{-0.90}$ (4/4)	$\mathbf{-0.75}$ (3/4)	-0.28 (0/4)
HF partial power	$\mathbf{+0.80}$ (4/4)	$+0.45$ (1/4)	$+0.15$ (0/4)

Table 5.3: BTF speed sensitivity, 0–$15\ \mathrm{barg}$.

5.5. Directional Repeatability

Each speed level contains anticlockwise and clockwise passes, revisiting the same commanded state within a recording. This tests both repeatability and the pooling assumption of section 5.4. Repeatability is quantified by the symmetric relative difference of equation (3.14) and classified with the thresholds of section 3.6: below $10\,\%$ repeatable, $10\,\%$ to $25\,\%$ marginal and above $25\,\%$ unreliable. Of the 270 pressure–speed–channel combinations, 89 contain at least 30 retained hits in both directions (24, 49 and 16 in the low, mid and high bands). Band-resolved results are given in table C.12.

Per-hit descriptors are generally repeatable. Median amplitude differs by at most $0.6\,\%$, spectral descriptors by $3\,\%$, and waveform RMS and peak by $4\,\%$. Energy and the two derived ratios are also

repeatable, counts, duration and rise time are marginal in their worst bands, and partial-power repeatability is band-dependent. Hit-rate is the exception, differing by $46\,\%$ to $81\,\%$ between directions per minute and by $65\,\%$ to $76\,\%$ per rotation; normalising by nominal rotation does not restore repeatability.

Timing error is a plausible contributor, because hit-rate depends on window duration whereas per-hit medians do not. Neither a central $60\,\%$ window nor a nominal one-rotation window resolves the disagreement, because the logged boundaries are shifted rather than merely over-wide. The disagreement is systematically signed: the first, anticlockwise pass records fewer hits, with a median count ratio of 0.61. Shifting all logged starts later by $1.5\,\mathrm{min}$ reduces the pooled median directional difference from $58.4\,\%$ to $25.2\,\%$ and moves the median ratio from 0.75 to 0.97. Splitting individual passes into halves, however, still gives large within-pass differences. Timing therefore explains the systematic bias but not the full variance. The fitted shift is retained only as a diagnostic: applying it to claim repeatability would be circular, and the observed lag is progressive rather than constant.

Value repeatability and relationship stability are distinct. Direction-specific correlations with speed and pressure were therefore computed for each feature and band, giving 114 paired comparisons. Signs differ in 39, but all reversals occur in weak relationships: no reversed mean exceeds 0.42 in magnitude. No direction-resolved coefficient reaches the 0.7 screen, because splitting the data halves the support, so the check shows only that reversals occur where direction is poorly determined. Median amplitude and waveform RMS retain the same sign between directions on the relevant low- and mid-band axes, with directional differences no greater than 0.086; their only reversal occurs for weak high-band pressure relationships.

Pooling is therefore supported for per-hit descriptors, particularly median amplitude and waveform RMS. Hit-rate does not support it: its pressure trend is directionally consistent in the pooled analysis but rests on a quantity with poor fixed-state reproducibility, and it is retained only as a descriptive activity indicator.

5.6. Operational Correction and Eligibility

The correction framework of Chapter 3 is applied to identify features with a stable pressure- and speed-dependent reference response. The complete mid-band grid contains 90 pressure–speed–channel combinations; 24 have fewer than 30 retained hits and one is empty. Within the primary 0–$15\ \mathrm{barg}$ range, 53 of 60 combinations meet the criterion, which is why the models are restricted to this range. Coverage is uneven, especially on channel 8 at $0\ \mathrm{barg}$, so models are evaluated separately by channel.

5.6.1. Correction Models

The observation unit is the aggregated pressure–speed–channel cell of section 5.2, pooled over rotation direction. For cell i, the feature value x_i is the median of the per-hit descriptor over the retained hits in that cell. For hit-rate per rotation, x_i is the retained hit count divided by the two nominal rotations of each speed level. Five features are modelled: hit-rate per rotation, median amplitude, waveform RMS, crest factor and kurtosis. Each is modelled separately on mid-band channels 2, 5 and 8. Cells below the 30-hit criterion are excluded, and a feature–channel combination is modelled only if at least three pressure levels and eight cells remain.

Let v_i denote the commanded rotation-speed magnitude of cell i in $\mathrm{deg/s}$, and p_i the gauge pressure p_g of the recording that contains it. Both predictors are standardised within each training fold:

$$v_i^* = \frac{v_i - \overline{v}_\mathcal{T}}{s_{v,\mathcal{T}}}, \qquad p_i^* = \frac{p_i - \overline{p}_\mathcal{T}}{s_{p,\mathcal{T}}}, \tag{5.5}$$

where $\mathcal{T}$ is the set of cells in the training fold, $\overline{v}_\mathcal{T}$ and $\overline{p}_\mathcal{T}$ are their mean speed and pressure, and $s_{v,\mathcal{T}}$ and $s_{p,\mathcal{T}}$ the corresponding population standard deviations. The same scaling parameters are applied to the held-out cells. The transformed feature value is

$$\tilde{x}_i = \begin{cases} \ln(x_i), & \text{Hit-rate per rotation or waveform RMS,} \\ x_i, & \text{amplitude, crest factor or kurtosis.} \end{cases} \tag{5.6}$$

Amplitude is not transformed because it is already expressed in decibels. Cells with $x_i \leq 0$ are excluded for the log-transformed features.

Two models are compared. The additive model (model A) is

$$\widehat{x}_i^{(A)} = \widehat{\beta}_0^{(A)} + \widehat{\beta}_v^{(A)} v_i^* + \widehat{\beta}_p^{(A)} p_i^*, \tag{5.7}$$

where $\widehat{x}_i^{(A)}$ is the predicted transformed feature value, $\widehat{\beta}_0^{(A)}$ the intercept, and $\widehat{\beta}_v^{(A)}$ and $\widehat{\beta}_p^{(A)}$ the speed and pressure coefficients. The hat denotes a least-squares estimate. Because the predictors are standardised, each coefficient gives the change in $\tilde{x}_i$ per standard deviation of its predictor, so the speed and pressure coefficients are directly comparable. The interaction model (model B) adds the product of the standardised predictors:

$$\widehat{x}_i^{(B)} = \widehat{\beta}_0^{(B)} + \widehat{\beta}_v^{(B)} v_i^* + \widehat{\beta}_p^{(B)} p_i^* + \widehat{\beta}_{vp}^{(B)} v_i^* p_i^*, \tag{5.8}$$

where $\widehat{\beta}_{vp}^{(B)}$ is the interaction coefficient. It allows the speed dependence to vary with pressure. Coefficients are estimated by unweighted ordinary least squares, so each cell carries equal weight irrespective of its hit count.

Both models are evaluated by leave-one-pressure-out cross-validation over 0, 5, 10 and $15\ \mathrm{barg}$. Each fold holds out all cells of one pressure recording. The scaling parameters and coefficients are estimated from the remaining three. This avoids the optimistic residual reduction obtained by evaluating only on fitting data. For each feature–channel combination, the model with the lower cross-validated root-mean-square error (RMSE) on the transformed scale is selected. The operational residual is the held-out prediction error,

$$r_i = \tilde{x}_i - \widehat{x}_i^{(-p_i)}, \tag{5.9}$$

where $\widehat{x}_i^{(-p_i)}$ is the prediction of the selected model fitted without the cells at pressure p_i. Residuals for hit-rate per rotation and waveform RMS are therefore on the natural-log scale.

Predictive adequacy is expressed as the cross-validated variance reduction, in per cent:

$$\Delta_{\mathrm{var}} = 100 \left[1 - \frac{\mathrm{Var}(r)}{\mathrm{Var}(\tilde{x})} \right], \tag{5.10}$$

where $\mathrm{Var}(r)$ is the population variance of the held-out residuals and $\mathrm{Var}(\tilde{x})$ that of the transformed feature values, both over the modelled cells of one feature–channel combination. A negative value means that the held-out residuals vary more than the uncorrected feature. Because $\mathrm{Var}(r)$ excludes the mean residual, systematic offset is checked separately by comparing the cross-validated RMSE with that of the training-fold median.

Of the 15 feature–channel combinations, nine retain the additive form (equation (5.7)) and six the interaction form (equation (5.8)), most often on channel 5. No quadratic pressure term is added. The non-monotonic response above $15\ \mathrm{barg}$ is treated as a limit of the model rather than fitted away.

5.6.2. Correction Performance and Eligibility

Median amplitude and waveform RMS provide the most consistent corrections. Their raw speed correlations are generally -0.70 to -0.88 across mid-band channels; cross-validated residual correlations fall to $+0.02$–$+0.25$. Variance reduction is $28\,\%$, $59\,\%$ and $25\,\%$ for amplitude on channels 2, 5 and 8, and $25\,\%$, $61\,\%$ and $19\,\%$ for waveform RMS. Improvement is therefore consistent but channel-dependent. Other features are less transferable. Hit-rate per rotation changes variance by -3, $+36$ and $+23\ \%$ across channels 2, 5 and 8, and crest factor by $+22$, $+45$ and $-8\ \%$. Kurtosis improves by $83\,\%$ on channel 5 but worsens by $15\,\%$ and $50\,\%$ on channels 2 and 8. A strong single-channel result is therefore insufficient for a general correction.

Models fitted over 0–$15\ \mathrm{barg}$ do not extrapolate to 20 and $25\ \mathrm{barg}$. Large residuals occur for both activity and amplitude-related features, because the measured response changes direction while the retained model is linear and because the upper-pressure recordings have much smaller retained populations. These residuals bound the valid correction range and are not treated as evidence of abnormal bearing condition.

Table 5.4 summarises the criteria for the mid band, which carries the correction. Association is the strongest mean Spearman coefficient on either axis, taken from the reduced sweep except for crest

factor and kurtosis, whose associations are strongest over the complete sweep (table C.9). Stability is judged from the sign-consistency fraction ϕ and the spread s_r of the underlying repeat coefficients, following section 3.6. All associated relationships have $\phi = 1.00$, meaning that every repeat has the same direction as the mean, and they are therefore marked stable; this holds even where an individual repeat falls below the association screen, as for amplitude ($s_r = 0.29$). Spectral descriptors, marked by a dash, fail the association screen and are therefore not assessed for stability. Repeatability is the worst-band median difference from table C.12. Predictive adequacy is the cross-validated variance reduction (equation (5.10)) on mid-band channels 2, 5 and 8; negative values indicate that the held-out residuals vary more than the uncorrected feature. Range coverage is 0–$15\ \mathrm{barg}$ for every row and is omitted as a column. Predictive adequacy distinguishes this dataset from the linear-bearing datasets, because the BTF provides both a reference condition and pressure recordings that can be held out during model fitting.

Feature (mid band)	Association (axis, $\bar{r}_s$)	Stab.	Repeat. (%)	Predictive adequacy (%)	Assessment
Amplitude (dB)	Speed -0.82	✓	0.6	$+28/+59/+25$	**Accepted**; residual carried forward
Waveform RMS	Speed -0.95	✓	2.9	$+25/+61/+19$	**Accepted**; residual carried forward
Crest factor	Press. -0.77	✓	3.3	$+22/+45/-8$	Candidate; fails predictive adequacy
Kurtosis	Press. -0.81	✓	9.0	$-15/+83/-50$	Candidate; fails predictive adequacy
Hit-rate (per minute)	Press. $+0.88$	✓	80.7	Not evaluated	Rejected at repeatability
Hit-rate (per rotation)	Press. $+0.88$	✓	76.0	$-3/+36/+23$	Rejected at repeatability
Energy	Speed -0.95	✓	8.8	Not modelled	Descriptive; associated and repeatable
Counts	Speed -0.94	✓	11.8	Not modelled	Descriptive; repeatability marginal
Duration	Press. $+0.80$	✓	11.9	Not modelled	Descriptive; repeatability marginal
Rise time	Press. $+0.80$	✓	14.6	Not modelled	Descriptive; repeatability marginal
Spectral descriptors	< 0.7	—	≤ 2.9	Not modelled	Repeatable but weakly associated
Partial powers	Speed ± 0.75	✓	0.2–21.8	Not modelled	Descriptive; repeatability band dependent

Table 5.4: BTF operational-relationship assessment.

Association alone therefore does not determine correctability. Every associated relationship is also directionally stable, so stability does not discriminate between features in this dataset; the decisive criteria are repeatability and predictive adequacy. Hit-rate is strongly associated but fails repeatability; spectral descriptors are highly repeatable but weakly associated; crest factor and kurtosis fail predictive adequacy. Only median amplitude and waveform RMS satisfy the complete criterion sequence: they are operationally associated, directionally stable within the supported range, repeatable to within $3\,\%$, adequately covered, and predictively adequate on held-out pressure recordings. Features not accepted remain as descriptive features.

5.7. Limitations

Independent operating states. The analysis rests on few independent operating states, and the correlation estimates are not independent of one another. In the reduced sweep, each pressure coefficient uses four operating points and each speed coefficient five. At $n = 4$ no Spearman coefficient can reach a two-sided $p = 0.05$, at $n = 5$ only $|r_s| = 1$ does, and at $n = 6$ approximately $|r_s| \geq 0.89$ is required, so the 0.7 screen denotes a trend rather than formal significance. Sign reversals are common across the underlying estimates; section C.8 should be consulted where a mean masks disagreement. The primary correction model likewise contains four pressure recordings, so each leave-one-pressure-out fold is fitted on only three pressure levels. This is conservative relative to hit-level fitting but remains a small-sample assessment.

Confounding. Pressure is confounded with the descending test order and may therefore absorb unmeasured changes in time, temperature, lubrication or sensor coupling. At 20 and $25\ \mathrm{barg}$, pressure

is also incompletely separated from speed: at $20\ \mathrm{barg}$, only $0.8\,\%$ of retained hits occur at the slowest level while $67.8\,\%$ occur in the two fastest.

Timing of the speed windows. Speed windows rely on a minute-resolution operator log and progressively lag the measured activity (section 5.5). The lag is small relative to the three slowest passes but comparable with the duration of the 4.5 and $6.0\ \mathrm{deg/s}$ passes, so results at the highest speeds are indicative. Neither duration rescaling nor AE-derived quiet intervals provided a defensible correction. About $22.7\,\%$ of hits fall in stationary intervals: $8.4\,\%$ in inter-pass pauses and $14.3\,\%$ after the final logged pass, concentrated in the final $6.0\ \mathrm{deg/s}$ clockwise window. Restricting the analysis to the three slowest speeds changes several feature-level conclusions, so the reported ranking applies to the full nominal speed grid.

Amplitude screening and windowing. The amplitude screen is absolute rather than noise-referenced and removes hits unevenly: the four lower-pressure recordings lose $22\,\%$ to $27\,\%$, while 20 and $25\ \mathrm{barg}$ lose $47.4\,\%$ and $42.5\,\%$, leaving only 350 hits at $25\ \mathrm{barg}$. The 30-hit windowing threshold is pragmatic and also removes observations unevenly. Channel 5 retains the most complete grid, so part of its stronger correction performance may reflect greater statistical support.

Sensors and processing. Channel 5 supplies $65.8\,\%$ of retained hits and almost all high-band hits come from channel 6. Source distribution, coupling and transmission-path effects cannot be separated, so high-band results are weakly constrained. Waveform decoding was not validated against the original Vallen software, and the partial powers use overlapping, incomplete passbands.

Scope of the correction. The correction is derived from a bearing without known damage. It characterises reference operational variability but does not establish that the same pressure–speed relationship persists with degradation. Residuals are operationally corrected measurements, not a validated damage index.

5.8. Chapter Summary

Across the full pressure sweep, scalar features show limited pressure sensitivity; several relationships strengthen only after 20 and $25\ \mathrm{barg}$ are removed. Speed has clearer effects, particularly on amplitude-related measures. Waveform crest factor and kurtosis show pressure dependence over the complete range but are sensitive to screening and do not generalise consistently across channels.

Association and repeatability rank the feature set differently. Hit-rate produces some of the largest coefficients but is the least repeatable measure, whereas spectral descriptors are highly repeatable but weakly associated. Median amplitude and waveform RMS are the only features that perform consistently across association, stability, repeatability, coverage and held-out predictive adequacy. Their operational residuals (equation (5.9)) are therefore carried forward.

The BTF dataset thus provides the thesis's full correction-eligibility demonstration for a bearing without known damage. Its accepted per-feature verdicts, rather than the BTF observations themselves, inform Chapter 6; the rigs do not share a compatible window-level observation unit for joint embedding. The correction is supported over 0–$15\ \mathrm{barg}$, and repeatability is demonstrated only through within-recording directional revisits.

6

Feature Engineering and Exploratory Modelling

This chapter develops the window-level AE feature representation used for exploratory multivariate analysis of the HiTeAM Phase 2 bearing data. Each mechanical window is represented by a vector of AE features. The resulting feature space is examined to establish how its dominant structure relates to operating state and contamination stage. TestBlock 2 provides an intact reference for operationally structured variation. TestBlock 3 contains an intact baseline followed by three successive contamination stages. PCA is used to examine the dominant linear covariance structure.

A first baseline-fitted residual representation is also examined. Its purpose is to establish whether structure associated with the measured operating state can be reduced before differences in the joint feature response are interpreted. These residual results are preliminary and are included only to establish the direction of the ongoing normalisation analysis. No classifier is fitted and no diagnostic accuracy is claimed.

6.1. Rationale and Scope

Earlier chapters characterised individual AE descriptors one at a time. This chapter considers them jointly to establish how the operational variation already identified appears when several descriptors are analysed simultaneously.

The analysis separates three roles. Load, speed and speed-profile phase are operating variables. The TestBlock 3 labels Baseline, Sample 1, Sample 2 and Sample 3 identify the successive contamination stages and serve as interpretation labels. AE activity and hit descriptors form the feature space.

The contamination stages are ordered by increasing iron content in the grease (table 4.10). Baseline is uncontaminated, Sample 1 and Sample 2 are intermediate levels at approximately $16\,\mathrm{mg\,kg^{-1}}$ and $800\,\mathrm{mg\,kg^{-1}}$, and Sample 3 is the highest level at approximately $40\,000\,\mathrm{mg\,kg^{-1}}$.

In the acquisition record these contamination stages are identified by the run label. Here, “run” does not denote an independent repetition under identical bearing conditions. It identifies a baseline and the three successive contamination stages, which were acquired sequentially. Contamination stage and acquisition order cannot be separated. These labels are referred to below as the *stage label*. Structure associated with them is described as stage-associated rather than as evidence that contamination alone produced the observed differences.

The speed programme ramps the commanded speed step as $0 \rightarrow 1 \rightarrow 2 \rightarrow 1 \rightarrow 0$ within each recording (section 4.3.2), and the five positions of this ramp are the speed-profile phases used here. Phases 1 and 2 are the ascending part at speed steps 0 and 1, phase 3 is the maximum-speed part at speed step 2, and phases 4 and 5 are the descending return through speed steps 1 and 0. Phases 1 and 5 share the lowest commanded speed and phases 2 and 4 the intermediate speed, but each pair

occurs at a different point in the programme, which is why speed-profile phase and nominal speed are treated as separate variables. The term speed-profile phase replaces the earlier “programme position” because the variable describes the dynamic phase of the prescribed speed programme rather than a physical position.

The analysis unit is a mechanical window rather than an individual hit. Hits within one operating segment are not independent observations of the operating state. Whole-file summaries are also too coarse because several files contain multiple load and speed states. Window-level aggregation yields comparable observations. Each observation is tied to a defined operating state and contains enough AE activity for stable feature estimates.

The chapter asks three questions. First, do intact data show systematic structure in the joint feature space under changes in operating state? Second, does TestBlock 3 contain additional structure associated with the stage label? Third, does a first baseline-fitted residualisation reduce structure associated with the measured operating state? The robustness of the result to feature-vector choice is examined in section 6.5.4.

PCA provides a linear representation of the dominant covariance structure, identifying the feature combinations that account for the largest variation between windows [14, 41]. It is not treated as a classifier or a causal model.

6.2. Window-Level Feature Construction

The feature-construction sequence is shown in figure 6.1. Raw AE records are assigned to reconstructed operating states. Mechanical cycles are then identified and screened, accepted cycles are grouped into fixed-length windows, and band-level features are aggregated within each window. One window is one observation. Each row of the resulting table corresponds to a mechanical window, and the columns contain its AE features.



Figure 6.1: TestBlock-3 feature construction, from individual AE records to window-level observations.

Operating states were reconstructed from file information, operator markers and SCADA position data. A valid operating block is a continuous interval of constant stage label, file, load step, speed step, speed-profile phase and segment part. Windows were not allowed to cross these boundaries. Mechanical cycles were identified as trough–peak–trough sequences in the measured position signal and retained only where the turning-point sequence and cycle duration were physically plausible.

Accepted cycles were grouped into non-overlapping windows of fixed length. The window length was selected before the dimensionality reduction, so that it was not chosen on the basis of the PCA result.

For window lengths of one to five cycles, the retained fractions were 100.0%, 96.3%, 94.2%, 92.3% and 83.3%, respectively. These fractions refer to accepted mechanical cycles that could be placed in complete non-overlapping windows within the operating-state boundaries. They do not refer to AE hits or files.

Four cycles were selected as a compromise between coverage and within-state feature stability. A longer window provides more AE observations per feature estimate but places fewer cycles in complete windows. At one cycle, the median coefficient of variation of mid-band hit rate is 0.583 and 9.7% of windows contain no mid-band hits. At four cycles these fall to 0.221 and 2.6%, while 92.3% of accepted cycles are retained. Five cycles retain only 83.3%.

This produced 313 four-cycle windows from 1356 accepted cycles, representing 1252 cycles, or 92.3%. The main analysis uses the 294 windows of the primary protocol: 73, 73, 76 and 72 for Baseline through Sample 3. The remaining 19 windows come from an additional repeat pass at load 1 during Sample 1 and are excluded, so that every stage is represented by the same test protocol.

A hit is assigned to window w when its timestamp falls within the window boundaries,

$$t_{\mathrm{start},w} \leq t_{\mathrm{hit}} < t_{\mathrm{end},w}, \tag{6.1}$$

where $t_{\mathrm{start},w}$ and $t_{\mathrm{end},w}$ are the temporal boundaries of window w and t_{hit} is the hit timestamp.

This assignment maps 4,300,914 of 5,383,227 hits (79.9%) to accepted windows. The unassigned fractions are 20.3%, 23.8%, 18.5% and 18.6% for Baseline through Sample 3, against 20.1% overall, so no stage is disproportionately excluded by the window construction.

Features were aggregated separately in the low, mid and high bands, following the acquisition-band definitions used in section 4.2.3. Activity is represented by hit count and normalised hit rate. Hit-conditional descriptors are summarised by medians and selected percentiles. Where a band contains no hits, activity variables are zero, while hit-conditional descriptors remain undefined because no event exists from which to calculate them.

A corresponding TestBlock 2 table was constructed as an intact reference, with windows formed within marker-defined speed dwell segments at the same four-cycle length, so that both blocks share one observation definition. TestBlock 2 produced 73 non-overlapping four-cycle windows from 318 accepted cycles. These windows represented 292 cycles, or 91.8% of the accepted-cycle population.

6.3. Feature Quality and Primary Vector

The exploratory feature vector is selected before PCA using three criteria: availability across the analysed windows, absence of known acquisition censoring, and avoidance of duplicate representations. Physical interpretability resolves cases the first three do not settle. Three quality-control findings determine the outcome.

First, 70 of the 294 primary windows contain no high-band hits. These are concentrated in Baseline and Sample 1, whereas every Sample 2 and Sample 3 window contains high-band activity. Conditional descriptors such as amplitude, energy or duration cannot be defined for a window without high-band events, and imputing them would create artificial event properties. High-band conditional descriptors are therefore not imputed.

Second, low-band duration is censored. The value $104\,857.6\,\mu\mathrm{s}$ appears as the median low-band duration in 126 of 294 primary windows (section 4.4.2). The repeated limiting value shows that low-band duration is not behaving as an unconstrained continuous descriptor. It is excluded before dimensionality reduction rather than treated as a physically meaningful extreme.

Third, the feature table is strongly redundant. The quality-control correlation analysis identifies 208 strongly correlated pairs, including exact rescalings and near-duplicate activity measures. Redundancy motivates a reduced primary vector, but correlation alone is not used as an automatic exclusion criterion. Mutually correlated descriptors are retained where they represent distinguishable physical aspects of AE behaviour.

The resulting primary vector, referred to as the low/mid core vector, contains 13 features (table 6.1). The low and mid bands each contribute hit rate as the activity measure, together with the median amplitude, rise time, energy, RMS and counts of the retained hits. Mid-band duration is also included, as it is not censored in that band. Hit count and hits per cycle are omitted: they are exact rescalings of each other, and they differ from hit rate only through the window duration, which is fixed in cycles and therefore varies with speed. On the log scale mid-band hit count and hit rate correlate at $r = 0.99$. Energy, RMS and counts are retained despite their mutual correlation because they describe different aspects of event strength. The vector therefore spans activity, event amplitude, event shape and event strength in the two bands that carry the bulk of the recorded activity.

Two sensitivity vectors test whether the main raw-to-residual result depends on this selection. Because PCA is determined by the covariance structure of the included features, changing the vector can change both the principal components and the variation associated with each label. The extended low/mid vector adds the hit counts and percentile descriptors that were omitted from the core vector as redundant. The high-band sensitivity vector adds high-band activity and presence to the core vector. High-band presence is used only as a sensitivity case because zero-hit windows are concentrated in the earlier stages, so presence itself carries stage information.

		Feature vector		
Feature family	Band	Core	Extended	Core + High-band
Hit rate	Low, mid	✓	✓	✓
Median amplitude, rise time, energy, RMS, counts	Low, mid	✓	✓	✓
Median duration	Mid	✓	✓	✓
Hit count, hits per cycle	Low, mid	✗	✓	✗
Percentile descriptors	Low, mid	✗	✓	✗
Activity and presence	High	✗	✗	✓
Median duration	Low	✗	✗	✗
Conditional descriptors	High	✗	✗	✗

Table 6.1: Feature families included in the low/mid core vector and the two sensitivity vectors.

6.4. Exploratory PCA Procedure

The analysis uses the 294 primary-protocol windows.

Windows with missing mid-band conditional descriptors are imputed before transformation. In Test-Block 3, the five affected primary windows are imputed from Baseline primary-protocol medians, so the imputation draws on the intact reference rather than on the later contamination stages. In TestBlock 2, the six affected windows are imputed from the TestBlock 2 median.

Non-negative non-dB descriptors are transformed as

$$\tilde{x}_{ij} = \log(1 + x_{ij}), \tag{6.2}$$

where x_{ij} denotes feature j for window i. Logarithmic transformation can reduce scale dependence and compress strongly right-skewed positive variables before PCA [59]. The addition of one keeps the transformation defined at zero. Energy and counts show particularly long upper ranges in the present data, while amplitudes in dB are left unchanged because the decibel scale is already logarithmic.

The transformed matrix is robust-scaled as

$$z_{ij}^{\mathrm{raw}} = \frac{\tilde{x}_{ij} - \mathrm{median}_{k\in\mathcal{P}}(\tilde{x}_{kj})}{\mathrm{IQR}_{k\in\mathcal{P}}(\tilde{x}_{kj})}, \tag{6.3}$$

where $\mathcal{P}$ denotes the primary windows. PCA is sensitive to differences in variable scale, so scaling is required to prevent features with larger numerical ranges from dominating the covariance structure [59]. Each feature is therefore centred on its primary-window median and divided by its interquartile range (IQR), defined as the difference between the 75th and 25th percentiles. The median and IQR are used rather than the mean and standard deviation because they are less sensitive to extreme observations [60].

Robust scaling equalises the interquartile range of each feature, not its variance. A feature with heavy tails can therefore retain a larger variance after scaling, and in these data mid-band energy has by far the largest variance of the scaled features, which is reflected in its leading loadings in both TestBlock 2 and raw TestBlock 3 PC1.

Before PCA, the IQR of each transformed feature column was checked explicitly. Columns with zero IQR cannot be divided by their IQR and were therefore assigned a unit scaling factor. This occurred only for the binary high-band presence feature in the high-band sensitivity vector; all columns in the primary 13-feature vector had non-zero IQR. After scaling, the feature matrix was checked for non-finite values (`NaN` or $\pm\infty$) before PCA was applied. The resulting PCA scores and loadings were checked in the same way and contained no non-finite values.

6.4.1. Preliminary Operational Residualisation

The residualisation follows the regression-based approach to operational and environmental normalisation established in SHM [48, 13, 47]. Residualisation replaces each feature by its departure from a

model of the expected value at the measured operating state, so that the remaining variation is what the operating variables do not explain. In this framework, measured response features are modelled as functions of environmental or operational covariates, and departures from the fitted reference response are then examined. Neumann et al. formulate this as a regression of each output on the covariates, in which the fitted regression function represents the conditional mean, and consider additive specifications with and without interaction terms [47].

Here, this framework is implemented separately for each AE feature using intact baseline windows only:

$$\tilde{x}_{ij} = \beta_{0j} + \sum_{l} \beta_{Lj,l} I(L_i = l) + \sum_{p} \beta_{Pj,p} I(P_i = p) + \varepsilon_{ij}, \qquad i \in \mathcal{B}, \tag{6.4}$$

where L_i is the load step of window i, P_i its speed-profile phase, and $\mathcal{B}$ the baseline windows. The function $I(\cdot)$ is an indicator function, equal to one when the stated condition is satisfied and zero otherwise; it therefore selects the load-level and speed-profile-phase offsets corresponding to window i. The term β_{0j} is the intercept for feature j. The coefficients $\beta_{Lj,l}$ and $\beta_{Pj,p}$ are the offsets for load level l and phase p.

Load step and speed-profile phase enter the model as categories rather than as numerical values. The model therefore estimates a separate mean response for each load step and each phase, without assuming that the AE response changes linearly between them. One level of each variable serves as the reference category, and the coefficients of the other levels give their offset from it.

The model is additive: the effect of a given load step is assumed to be the same in every phase. Regression-based normalisation can also include interaction terms, which would allow the load effect to differ between phases [48, 47]. Such terms require baseline observations at every load–phase combination. All load steps and phases appear in the baseline, but the load 3/phase 4 and load 3/phase 5 combinations are absent, so an interaction model cannot be estimated for them. More generally, the number of reference states required grows rapidly as conditioning terms are added [48]. Predictions for these two combinations therefore rest on the additive structure of the model rather than on observed baseline data, and have weaker support.

Following the same regression-based normalisation principle, the departure from the fitted baseline response is calculated as a prediction residual [13, 47]:

$$r_{ij} = \tilde{x}_{ij} - \hat{x}_{ij}(L_i, P_i), \qquad z_{ij}^{\mathrm{res}} = \frac{r_{ij} - \mathrm{median}_{k \in \mathcal{B}}(r_{kj})}{\mathrm{IQR}_{k \in \mathcal{B}}(r_{kj})}, \tag{6.5}$$

where r_{ij} is the departure of the observed value from the baseline prediction at the same measured operating state. The residuals are subsequently centred and scaled on the baseline residual distribution. Median and IQR are used for this scaling to reduce the influence of individual extreme observations on the centring and scale estimates [60].

The residual space represents variation not explained by the fitted baseline load/phase mean response. It is not a validated operational normalisation. The model captures only additive mean shifts in the two measured operating variables, so load–phase interactions, nonlinear effects within a category, operating-dependent variance [47] and the influence of unmeasured variables may all remain in the residuals. The residual analysis here is therefore preliminary and shows only the first effect of removing the fitted baseline mean structure.

6.4.2. Retention and Label Association

For this exploratory analysis, a cumulative explained-variance criterion of 90% is adopted. Cumulative variance thresholds provide a common criterion for selecting the number of principal components, with values between 70% and 90% often used depending on the application [59]. The smallest number of principal components reaching the 90% threshold is retained for interpretation. The remaining components are not used in the group-association summaries.

PCA reorganises variance rather than removing it, so the reduction lies in this retention step and not in the decomposition itself.

Label association in the retained PCA score space is summarised by R^2_{group}, defined as the proportion of total score-space variation associated with between-group centroid differences. This follows the additive geometric partitioning of multivariate Euclidean variation [61].

$$R^2_{\text{group}} = \frac{\sum_g n_g \left\|\bar{\mathbf{z}}_g - \bar{\mathbf{z}}\right\|^2}{\sum_i \left\|\mathbf{z}_i - \bar{\mathbf{z}}\right\|^2}, \tag{6.6}$$

where $\mathbf{z}_i$ is the retained PCA score vector for window i, $\bar{\mathbf{z}}_g$ the centroid of group g, and $\bar{\mathbf{z}}$ the overall centroid. The numerator is the between-group sum of squares and the denominator is the total sum of squares in the retained PCA representation. Thus, R^2_{group} describes the fraction of retained score-space variation associated with the grouping variable. It is used descriptively and not as a significance test or classification metric.

In the results, the same measure is calculated separately for load step, nominal speed, speed-profile phase and contamination stage. A larger R^2_{group} means that a greater fraction of the retained PCA variation is associated with differences between the groups defined by that label. A value close to zero indicates little organisation of the retained score space by that label. The measure does not indicate the direction of the relationship.

6.5. Exploratory Results

6.5.1. TestBlock-2 Reference

In the raw low/mid core space, TestBlock 2 PC1 explains 73.5% of total variance and the first three components 93.3%.

Speed-profile phase has the strongest association with the retained scores ($R^2_{\text{group}} = 0.360$), followed by load (0.158) and nominal speed (0.136). The intact bearing already shows systematic structure in its joint feature response under operating changes alone (figure 6.2). Separation in PCA space cannot by itself be read as evidence of degradation.

The load result must be treated separately from the speed-related ones. The acquisition threshold was retuned during the TestBlock 2 load sequence (section 4.3.3) and changed systematically with load. Because a threshold change alters which hits are recorded, and with them retained amplitude, duration, counts and activity, the effects of physical load and threshold retuning cannot be separated in these data. The load association is reported only as a description of the observed organisation, not as evidence of a physical load effect.

Component	Variance	Cumulative	Largest absolute loadings
PC1	73.5%	73.5%	Mid-band energy ($+0.742$), mid-band amplitude ($+0.439$), mid-band duration ($+0.416$), mid-band counts ($+0.250$)
PC2	13.9%	87.4%	Low-band energy ($+0.764$), low-band counts ($+0.423$), low-band amplitude ($+0.203$), low-band RMS (-0.198)
PC3	5.9%	93.3%	Mid-band rise time ($+0.668$), mid-band duration ($+0.348$), low-band RMS ($+0.309$), mid-band hit rate ($+0.286$)

Table 6.2: TestBlock-2 retained principal components and dominant loadings. The four largest absolute loadings are listed for each component.

Three components satisfy the 90% retention criterion, and table 6.2 lists them with their four largest loadings [59]. Each component is a weighted combination of the scaled features, and the loadings are these weights. A window's score on a component is the sum of its scaled feature values multiplied by the loadings. The squared loadings of a component sum to one, so a squared loading gives the share of the component direction contributed by that feature. Mid-band energy, with a loading of 0.742, accounts for 55% of PC1. Features with the same sign increase together along a component, and features with opposite signs change in opposite directions: low-band RMS has a negative loading on

PC2 and therefore decreases as low-band energy and counts increase along that direction. The overall sign of a component is arbitrary, so only relative signs within a component are interpreted.

The variance column gives the fraction of the total variation of the scaled features that lies along each component, and the cumulative column gives the fraction retained up to that component. A larger percentage is not better in itself; it indicates that more of the variation is concentrated in that direction. The loadings identify which features define each retained direction, whereas R^2_{group} describes how strongly a label organises the retained space as a whole. Because R^2_{group} is calculated over the retained components jointly, the score plots and loadings are used to see along which directions the groups differ.

The three retained components separate by band and by descriptor type. PC1 loads on mid-band energy, amplitude and duration together, and is read as a mid-band event-strength direction. PC2 is the corresponding low-band direction, dominated by low-band energy and counts. PC3 is led by mid-band rise time and duration and therefore reflects event shape rather than event magnitude. Although TestBlock 2 activity is overwhelmingly concentrated in the low band, mid-band activity is present in 67 of the 73 windows. Six windows contain no mid-band hits and their conditional descriptors are imputed. The median mid-band hit count is 26 per window, with an interquartile range of 8–37. PCA is applied to robust-scaled window-level features and does not weight columns by the number of underlying hits. Mid-band features can contribute strongly despite the smaller mid-band event population.

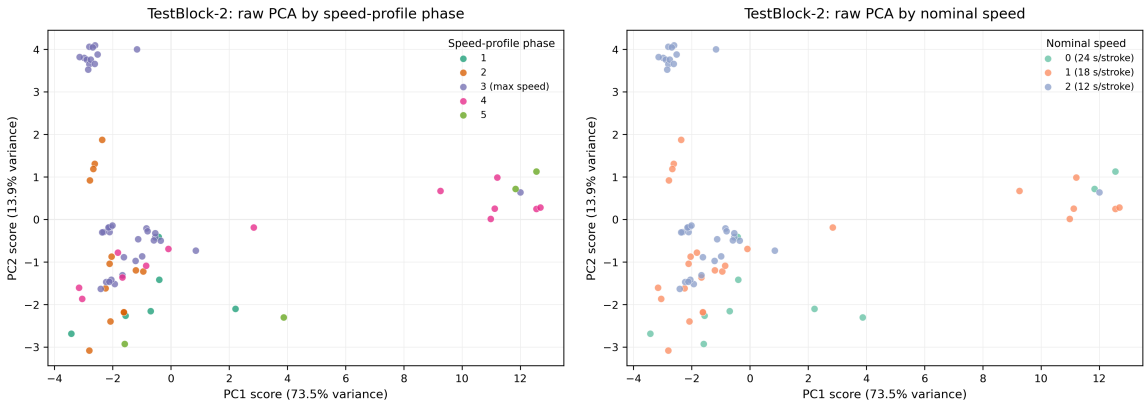


Figure 6.2: TestBlock-2 raw PCA, coloured by speed-profile phase (left) and nominal speed (right). Phase 3 is the maximum-speed phase; phases 1 and 5 are the slowest.

Residualising TestBlock 2 on the same observations used for fitting removes the fitted load/phase group means by construction. A reduction in operational grouping is expected on the fitting data and does not show that the correction would generalise. These results are treated only as a method check.

6.5.2. TestBlock-3 Raw and Preliminary Residual Structure

In the raw low/mid core space, four components reach the 90% retention criterion and together explain 91.2% of total variance (table 6.3). The retained score space is associated more strongly with nominal speed and speed-profile phase than with load. The speed and phase values of $R^2_{\text{group}} = 0.334$ and 0.353 mean that roughly a third of the variation in the retained score space lies between the speed groups or between the phases, whereas load accounts for only 4.3% ($R^2_{\text{group}} = 0.043$). Stage accounts for 26.9% ($R^2_{\text{group}} = 0.269$), so in the raw space operating state organises the feature space more strongly than contamination stage.

After baseline-fitted residualisation, four components explain 92.6% of the residual variance. Association with load, speed and phase falls to 0.010, 0.060 and 0.073, while stage association rises to 0.468. Residualisation reduces a substantial part of the structure associated with the measured operating variables and leaves stronger differences between stages. Because the stages were acquired sequentially, this remaining separation cannot be attributed to contamination alone.

Figure 6.3 shows the change directly. In the raw space, nominal speed forms a clear gradient across the first two components, so operating state contributes strongly to the dominant covariance structure.

Space	PCs	Cum. var.	Load R^2_{group}	Speed R^2_{group}	Phase R^2_{group}	Stage R^2_{group}
Raw	4	0.912	0.043	0.334	0.353	0.269
Residual	4	0.926	0.010	0.060	0.073	0.468

Table 6.3: TestBlock-3 PCA organisation, raw and residual.

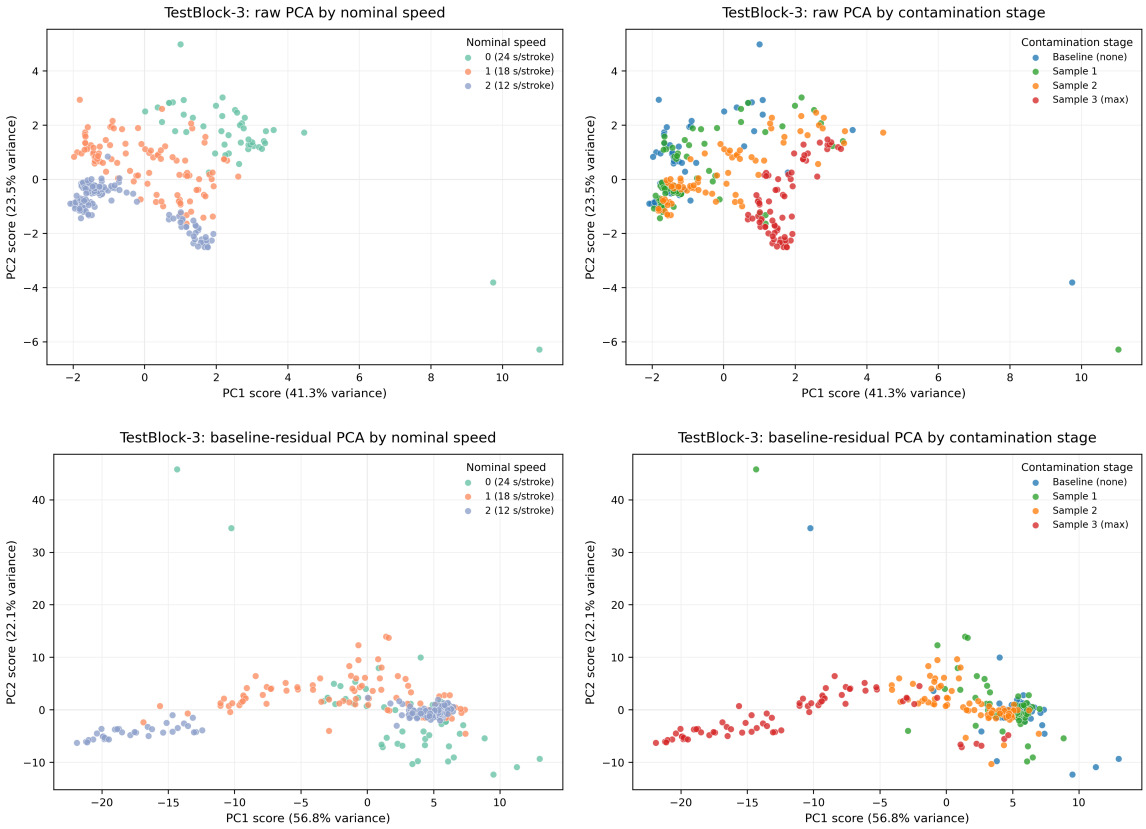


Figure 6.3: TestBlock-3 PCA, raw (top) and residual (bottom), coloured by nominal speed (left) and contamination stage (right).

After residualisation the speed groups overlap substantially more, consistent with the fall in the speed-associated R^2_{group}.

Two isolated points in figure 6.3 lie well outside the main cluster in both the raw and residual PCA score spaces. One belongs to Baseline and the other to Sample 1, and both occur during speed-profile phase 5. Inspection of the underlying AE hits shows that these windows contain high-amplitude broad-band bursts that dominate the sparse mid-band event population. Their mid-band median amplitude and energy are therefore determined by the bursts rather than by the more typical mid-band activity. The two windows are retained because they are valid measured observations. No acquisition or processing fault was identified for them, and removing windows because of their position in the PCA result would mean selecting data on the basis of the outcome. The robust centring and scaling limit their influence on the feature scaling, although they still contribute noticeably to residual PC2 (46% of its variance), which should be kept in mind when reading that component in table 6.4.

The stage-coloured residual space instead shows a pronounced displacement of Sample 3 along PC1, while Baseline, Sample 1 and much of Sample 2 remain closer together. The dominant organisation shifts from the measured operating variables towards stage-associated separation.

The residual PCA also suggests that the response across contamination stages may not be gradual. Sample 3, the highest contamination level, is displaced considerably further than the earlier stages. This is consistent with a nonlinear or threshold-like change in the joint feature response. It does not establish a contamination threshold because only four stages are available and contamination level changes together with acquisition order.

6.5.3. Principal-Component Loadings

Table 6.4 lists the retained components in both spaces with their dominant loadings, and is read in the same way as table 6.2. The raw and residual variance percentages are fractions of different totals: the larger share of residual PC1 indicates that the residual variation is more concentrated in one direction, not that it is larger. The leading raw directions do not correspond to single descriptors: each combines variation from several feature families and from both bands. Raw PC1 and PC2 together account for 64.8% of total variance and mix activity, amplitude and event-strength descriptors across the low and mid bands. Raw PC3 (20.4%) is led by mid-band duration and energy, and raw PC4 (6.0%) by mid-band RMS and mid-band duration. The event-shape descriptors, mid-band duration and low-band rise time, appear among the largest loadings of PC3 and PC4 but not of the two leading components.

Residualisation changes the structure as well as reducing the operating-state association. Residual PC1 alone explains 56.8% of the remaining variance and contrasts low-band event magnitude against mid-band amplitude, while residual PC2 is almost entirely mid-band. The residual representation is therefore more concentrated and more clearly separated by band than the raw one.

Two lower-variance residual directions are worth noting for later discussion. Residual PC3 (8.3%) is dominated by a single feature, low-band acquisition RMS ($+0.743$), whose loading exceeds any individual loading in residual PC1. Acquisition RMS describes the background level preceding the threshold crossing rather than the detected transient (section 3.4), so this component is read as a background-level direction that survives the baseline fit. Residual PC4 (5.4%) opposes mid-band hit rate and low-band energy against low-band RMS, and is the only retained residual component in which activity and background move in opposite directions. Neither is interpreted further here.

This interpretation remains descriptive. Because the stage label also carries acquisition order, the loadings identify which features contribute to the observed stage-associated structure but do not establish its physical cause.

6.5.4. Feature-Vector Sensitivity

The raw-versus-residual pattern also holds for both sensitivity vectors (table 6.5). The table repeats the PCA for each feature vector in the raw and the residual space and reports the R^2_{group} values for speed, phase and stage, with the low/mid core vector included for comparison. In every vector, speed and phase association fall after residualisation, while stage association rises. For the extended vector, speed and phase association fall from 0.386 and 0.404 to 0.056 and 0.075, and stage association rises from 0.206 to 0.549. The high-band sensitivity vector shows the same change, with stage association

Space	Component	Variance	Cumulative	Largest absolute loadings
Raw	PC1	41.3%	41.3%	Mid-band energy ($+0.519$), low-band hit rate ($+0.378$), mid-band counts ($+0.338$), low-band amplitude (-0.300)
	PC2	23.5%	64.8%	Low-band RMS ($+0.529$), mid-band hit rate (-0.423), low-band amplitude (-0.378), mid-band energy (-0.335)
	PC3	20.4%	85.2%	Mid-band duration ($+0.550$), mid-band energy ($+0.438$), mid-band hit rate (-0.386), low-band rise time ($+0.367$)
	PC4	6.0%	91.2%	Mid-band RMS ($+0.469$), mid-band duration ($+0.466$), low-band hit rate ($+0.382$), low-band rise time (-0.370)
Residual	PC1	56.8%	56.8%	Low-band energy ($+0.451$), low-band counts ($+0.427$), low-band RMS ($+0.426$), mid-band amplitude (-0.347)
	PC2	22.1%	78.9%	Mid-band energy ($+0.709$), mid-band amplitude ($+0.346$), mid-band duration ($+0.318$), mid-band counts ($+0.309$)
	PC3	8.3%	87.2%	Low-band RMS ($+0.743$), mid-band hit rate ($+0.437$), mid-band rise time (-0.234), mid-band duration (-0.223)
	PC4	5.4%	92.6%	Mid-band hit rate ($+0.459$), low-band energy ($+0.438$), low-band RMS (-0.368), low-band counts ($+0.355$)

Table 6.4: TestBlock-3 retained principal components and dominant loadings in the raw and residual low/mid core spaces. The four largest absolute loadings are listed for each component.

rising from 0.352 to 0.546. Its raw stage association is already higher than for the other two vectors because high-band presence is concentrated in the later stages (section 6.3). The result is not confined to the particular variables selected for the low/mid core vector.

Feature vector	Space	Speed R^2_{group}	Phase R^2_{group}	Stage R^2_{group}
Low/mid core	Raw	0.334	0.353	0.269
Low/mid core	Residual	0.060	0.073	0.468
Extended low/mid	Raw	0.386	0.404	0.206
Extended low/mid	Residual	0.056	0.075	0.549
High-band sensitivity	Raw	0.278	0.294	0.352
High-band sensitivity	Residual	0.050	0.061	0.546

Table 6.5: PCA feature-vector sensitivity.

6.5.5. Preliminary Residual Checks

The residual result is not yet a completed normalisation analysis. Two checks show where further work is required.

First, five windows had missing mid-band values that were filled in by imputation. Removing these windows hardly changes the raw PCA result: the R^2_{group} values for load, speed, phase and stage all change by less than 0.01. The residual result is less stable. Without the five windows, the remaining phase association rises from 0.073 to 0.253 and the speed association from 0.060 to 0.129. Stage association falls from 0.468 to 0.385, but is still the largest. None of the five windows is at load 3, so this change is not caused by the load–phase combinations missing from the baseline. It comes from phase 5, where

the baseline contains only a few windows. How much operational structure the residualisation appears to remove therefore depends on a small number of baseline windows at phase 5.

Second, feature-level residuals retain some phase dependence. The largest residual phase R^2_{group} values occur in mid-band hit rate (0.144), counts (0.132), duration (0.128), RMS (0.111) and energy (0.098). Residualisation reduces, but does not remove, measured operational structure.

These observations motivate a fuller assessment of model form, support and validation before the residual representation is treated as an operational normalisation.

6.6. Bounded Interpretation

The exploratory PCA establishes that operating state is a dominant source of structure in the joint AE feature space. This holds even for the TestBlock 2 bearing without known damage, so changes in operating state alone can produce systematic organisation of the feature space. Structure in a raw AE feature space can therefore not be read as condition-related without first accounting for the operating state in which it was measured.

The results also show that this operational structure can be reduced explicitly. A residualisation fitted on the intact baseline, using only the measured load and speed-profile phase, substantially lowers the association with the operating variables in TestBlock 3, and contamination stage becomes the dominant organising factor of the retained score space. This change holds across all three feature vectors, so it does not depend on the particular features selected. Because the stages were acquired sequentially, the remaining structure is interpreted as departure from the fitted intact-baseline response rather than as a direct measure of contamination or degradation.

The preliminary checks identify where the approach can be strengthened. Residual phase dependence remains, and the extent of the operational reduction depends on sparse baseline support at phase 5. Both point to specific improvements, namely denser baseline coverage and a model that can represent operating-dependent effects, rather than to a limitation of the approach itself. The chapter therefore provides the first demonstration that explicit operational residualisation can change the dominant organisation of an AE feature space, and defines what is needed to develop it into a validated normalisation.

6.7. Chapter Summary

A reproducible four-cycle feature table was constructed for TestBlock 3. The four-cycle length retains 92.3% of accepted mechanical cycles while substantially improving mid-band hit-rate stability and reducing mid-band zero-hit windows relative to single-cycle observations.

The table was reduced to the conservative low/mid core vector on grounds of availability, censoring, redundancy and physical interpretability. The main analysis uses 294 windows and excludes censored low-band duration and undefined high-band conditional descriptors.

TestBlock 2 shows that substantial PCA structure can arise from variation in operating state without a known change in bearing condition. TestBlock 3 shows strong speed- and phase-associated structure together with differences between contamination stages.

A first baseline-fitted residual representation reduces the measured association with load, speed and phase in TestBlock 3 while leaving stronger stage separation. This is a preliminary result rather than a validated normalisation. Residual phase dependence and sparse baseline support remain to be examined before the correction stage is complete. The chapter establishes the exploratory structure of the joint AE feature response and the first residualisation result on which the subsequent normalisation analysis builds.

7

Discussion

This chapter interprets chapters 4 to 6 against the three research sub-questions. Three findings are linked. AE features are strongly conditioned by operating state. Only a subset of operational relationships is suitable for correction. Reducing measured operational structure changes the dominant organisation of the multivariate feature space. The final sections translate these findings into monitoring implications and define the limits of the evidence.

7.1. Operational Influence on AE Features

7.1.1. Load and Pressure Dependence

Load is a first-order influence wherever it can be varied cleanly, but the response is a property of the configuration rather than of the feature. Phase 4 shows strong load coupling in high-band hit-rate, mid-band waveform RMS and mid-band waveform peak, at $\bar{r}_s = +0.89$, $+0.92$ and $+0.91$, yet no feature is load-associated in all three bands. These relationships describe a heavily worn bearing and should not be transferred to an undamaged contact.

The full-scale BTF response is less clear, and the reason matters more than the coefficients. Across 0 to $25\,\mathrm{barg}$ no quantity and band combination reaches the association threshold on average, whereas restricting the range to 0 to $15\,\mathrm{barg}$ gives hit rate correlations of $+0.88$ to $+0.96$. hit-rate rises towards 10 to $15\,\mathrm{barg}$ and then falls, so the weak full range coefficient reflects a response that changes direction rather than insensitivity to pressure. This is a limitation of a rank-order screen rather than a property of the bearing. Chapter 5 therefore tabulates the reduced sweep and reports the complete sweep in section C.8, quoting it wherever the two differ. The stronger reduced-range coefficients describe behaviour within the supported domain and are not improved estimates of the full-range response.

Pressure changes event occurrence more than event magnitude at BTF. hit-rate spans approximately two orders of magnitude, whereas median amplitude changes by at most $2.9\,\mathrm{dB}$. This is compatible with a reference bearing in which operating state changes AE activity without a comparable change in per-event severity. The campaign contains no damaged full-scale comparator, so that interpretation cannot be tested. The change above $15\,\mathrm{barg}$ also remains unresolved. The upper pressure recordings are sparse, pressure was adjusted manually, flange weeping was reported, and pressure is confounded with test order.

Several physical explanations are plausible. The 20 and $25\,\mathrm{barg}$ recordings were acquired first, when the bearing and grease may have been colder, and a higher lubricant viscosity could then have increased film separation and reduced asperity-related AE activity. The larger hydraulic separating force may also have changed the internal clearance or the load distribution within the swivel, altering which roller contacts are active. A third possibility is that the combined effect of pressure, speed, viscosity and contact conditions shifted the bearing into a different lubrication regime. These mechanisms cannot be distinguished from the present measurements. A repeated pressure sweep would need a balanced test order, calibrated pressure, measurements of bearing temperature and of friction or torque, and

internal displacement where access permits.

7.1.2. Speed and Exposure

Speed affects event rate and event magnitude differently, so a speed change cannot be described as simply raising or lowering AE level. In the Phase 4 constant load comparison, high-band hit-rate increases by a factor of 1.95 while high-band RMS and peak amplitude fall by factors of 2.96 and 3.39. The same contrast appears at BTF, where amplitude-related features fall with speed while time-based hit-rate rises.

The BTF results also show that the exposure definition is part of the feature definition. Mid band hit-rate increases with speed when expressed per minute, at $r_s \approx +0.88$, and decreases when expressed per nominal rotation. Time based hit-rate therefore mixes source activity with mechanical exposure.

Direction affects AE activity more than the properties of individual events. At BTF, hit-rate differs by 46 to 81 per cent between anticlockwise and clockwise visits to the same commanded state, whereas most per-hit descriptors differ by only a few per cent. Part of this disagreement comes from the operator log alignment, since shifting the logged boundaries by approximately $1.5\,\mathrm{min}$ reduces the median directional difference substantially. Large within pass variations remain, so timing uncertainty does not explain the whole effect.

A physical directional influence therefore cannot be excluded. Reversal may change the order of contact with local surface features, grease redistribution, cage behaviour or other direction-dependent contact conditions. The experiment cannot separate these mechanisms, because the anticlockwise pass always preceded the clockwise pass and direction is confounded with order of visit. The strong repeatability of the per-hit descriptors suggests that any physical directional effect acts on the occurrence of AE events rather than on their magnitude or shape. A repeated experiment should alternate the order of rotation, segment on direct angular position measurements, and include repeated passes in each direction at the same operating state.

7.1.3. Lubrication and Contamination

TestBlock 3 shows that contamination changes the operational relationship itself, which is the more consequential finding for correction. The low-band load coupling of acquisition RMS changes from -0.93 in the clean baseline to -0.07 at the highest contamination stage. Energy, counts and duration weaken similarly, while median amplitude and the derived ratios remain more stable. A correction fitted under one lubricant condition can therefore become inaccurate once that condition changes.

Contamination also changes the AE event population at comparable operating states. hit-rate approximately doubles, mean duration approximately halves, amplitude changes little, and activity increases strongly in the mid and high-bands. The later stages therefore contain more and shorter events rather than larger ones.

This behaviour cannot be transferred directly from earlier contamination studies. Miettinen and Andersson showed that AE activity is sensitive to grease contamination and reported an influence of particle size on the measured AE level [46]. The present experiment differs in particle size, concentration, bearing geometry and contamination procedure, since TestBlock 3 uses steel and silicon carbide particles up to approximately $500\,\mu\mathrm{m}$ introduced sequentially without cleaning between stages. The increase in event rate with little change in amplitude therefore represents a different population response. The mechanism cannot be established from the available measurements. What the result does show is that contamination signatures depend on configuration and should not be transferred between bearing systems without verification.

7.1.4. Feature, Band and Measurement Dependence

The datasets do not support a universal AE feature or frequency band. Amplitude related quantities are the most consistently responsive to operating state and the most reliable correction candidates. Hit and count features can be strongly associated but are sensitive to exposure definition, threshold and repeatability. Spectral descriptors are often highly reproducible yet weakly or inconsistently coupled. Reproducibility, operational sensitivity and correctability are therefore distinct properties, and a feature that scores well on one of them cannot be assumed to score well on the others.

Frequency band changes the interpretation in the same way. The same descriptor can carry different or opposite relationships between bands, and the dominant band differs between datasets. The low-band dominates much of Phase 2, the high-band gives the clearest Phase 4 load response, and the mid-band dominates BTF. Band usefulness is specific to the installation rather than intrinsic to the feature.

The measurement configuration matters as much as either. BTF activity is strongly imbalanced across channels, and TestBlock 2 shows that threshold retuning can create an apparent operational relationship by truncation alone. SNR filtering and amplitude screening also change the retained event population and, in some cases, the reported relationship. Sensor path, threshold, screening rule and retained population are therefore part of the measurement definition rather than incidental settings.

7.2. From Operational Association to Operational Correction

7.2.1. Eligibility and Held Out Correction at BTF

A strong operational association is necessary but not sufficient for correction, and each further criterion proved decisive for at least one feature that an association screen alone would have accepted. TestBlock 2 rise time is strongly associated with speed but poorly repeatable. TestBlock 3 RMS, energy and counts lose or reverse load coupling as contamination progresses. At BTF, several waveform-shape descriptors are strongly associated with pressure but fail held-out predictive adequacy across nominally equivalent channels.

Median amplitude and waveform RMS are the only BTF quantities that satisfy association, stability, repeatability, range coverage and held-out predictive adequacy together. Under leave-one-pressure-out validation both reduce residual speed association from approximately -0.8 to near zero and reduce variance by roughly 20 to 60 per cent across the three mid band channels. Correction targets should therefore be selected by predictive consistency rather than by association strength.

The supported correction domain of 0 to $15\,\mathrm{barg}$ follows from model support rather than from the behaviour of the excluded states. Within this range the retained populations sustain the pressure–speed medians, whereas above it they collapse, pressure and speed are incompletely separated, and the measured response changes direction so that a linear model fitted below the limit cannot represent it. The large residuals at 20 and $25\,\mathrm{barg}$ therefore bound the valid correction range. They are not evidence of abnormal bearing condition.

7.2.2. Changing Relationships and Limits of Residualisation

Correction assumes that the reference operational relationship remains relevant. TestBlock 3 shows that this assumption can fail, since contamination changes the load dependence of several features while the measured operating variables stay the same. Residuals can then contain both condition information and error from an outdated operational model.

Residualisation also removes only the fitted mean response. Residual variance, covariance, nonlinear dependence and unmeasured influences may remain, so corrected does not mean free of operational effects. The low-order models used here favour interpretability and generalisation from limited data, but they cannot represent all observed structure, including the unimodal BTF pressure response.

The comparison is asymmetric in one respect that should be stated. The reference model is fitted on baseline windows, so baseline residuals lie within the fitting set while the contamination residuals lie outside it. Residuals within the fitting set are systematically too small, which makes the baseline cloud appear more compact than it is and inflates the apparent separation. The effect is bounded here. Eight parameters are fitted on 73 baseline windows, giving a residual scale optimism of approximately 6 per cent, which is small relative to the displacement of the final contamination stage and does not account for it.

The TestBlock 2 case is the one place where the analysis would have been circular. Residualising those windows on the same observations used for fitting removes the fitted group means by construction. Chapter 6 therefore treats that result as a method check rather than as evidence, and no conclusion rests on it.

Two properties keep the TestBlock 3 result free of the same objection. The residual model never saw the

contamination stages, because its coefficients are fitted on baseline windows only. PCA is also unsupervised, so it maximises variance rather than class separation and receives no stage labels. Structure that appears across stages in the residual score space is therefore a property of the covariance structure rather than something the procedure imposed. A supervised projection would not permit the same reading.

7.2.3. Why Correction Models Do Not Transfer

Chapter 5 argued that the datasets should be analysed separately, because a joint PCA representation would risk separating rigs rather than separating operation from condition. That argument rested on expectation rather than evidence, and a check performed on the most favourable available pair supports it. TestBlock 2 and TestBlock 3 share a rig, a cycle definition, an observation unit and a set of conditioning variables, yet fitting the operational model on the first and applying it to the second degraded the representation. It raised load association well above its raw value and lowered stage association below it. The TestBlock 2 coefficients are confounded by the threshold retuning discussed in section 7.1.4, so subtracting them imposes that pattern on TestBlock 3, and the two residual distributions differ too much in spread for the TestBlock 2 scaling to weight the feature space correctly. Neither defect is visible in the quality of the fit on TestBlock 2 alone, which is the practical warning.

The BTF reference could not have supplied the operational model of chapter 6 for a stronger reason. The four-cycle observation unit does not exist at a continuously rotating swivel bearing, and the conditioning variables do not overlap, since the BTF model is a function of applied pressure and angular speed whereas the TestBlock 3 model is a function of load step and speed-profile phase. The two design matrices share no columns, so BTF coefficients are undefined for TestBlock 3 rather than merely inaccurate.

BTF instead contributes the procedure and the per-feature verdicts. It is the only dataset in this work where the complete eligibility sequence could be evaluated end to end, because independent pressure recordings allow predictive adequacy to be assessed by leave-one-pressure-out validation. TestBlock 3 cannot support that test, since it has no pressure axis and contamination stage is confounded with acquisition order. Its feature vector is broader and is used for exploratory multivariate analysis rather than as a set of validated correction targets. The two datasets therefore answer different questions. BTF establishes which features satisfy the complete correction criteria and whether the correction generalises to held-out operating states. TestBlock 3 examines how a baseline-fitted residual representation behaves across successive stages.

7.3. Separating Operational and Stage-Associated Variability

7.3.1. Raw and Residual Multivariate Structure

The pooled association measures reported in chapter 6 are re-examined here at component-level and within matched load–phase combinations, because the conclusion that follows from the pooled values alone does not survive that examination.

The raw TestBlock 3 PCA is organised mainly by operating state. Nominal speed and speed-profile phase reach $R^2_{\mathrm{group}} = 0.334$ and 0.353, against 0.043 for load and 0.269 for stage. TestBlock 2 gives the same warning in an intact bearing, where speed-profile phase reaches 0.360 with no contamination sequence present. Raw multivariate structure can therefore not be read as condition-related without first examining organisation by operating state.

After baseline-fitted residualisation, load, speed and phase associations fall to 0.010, 0.060 and 0.073, while stage association rises to 0.468. Taken at face value this is a reduction of 77 to 82 per cent in measured operational organisation. Two qualifications weaken that reading, and both follow from where in the score space the association sits rather than from the pooled value.

The first is that the pooled measure is variance weighted, since the multivariate R^2_{group} of equation (6.6) sums squared score distances across the retained components. Decomposing the residual phase association by component reproduces the pooled value and shows that it is unevenly distributed. Residual PC1 carries 56.8 per cent of the variance but a component-level phase association of only 0.021, whereas PC3 and PC4 carry 8.3 and 5.4 per cent of the variance at 0.174 and 0.183, which are compa-

rable to raw-space levels. The same decomposition attributes 94 per cent of the stage association to PC1. Speed-profile phase structure has therefore been displaced into directions that contribute little to the total sum of squares rather than eliminated, and the residual space is better described as one stage-associated direction alongside operationally organised minor directions.

The second qualification is that the stage displacement is itself operating dependent. The matched load–phase comparison contrasts each stage against the baseline median within the same load step and speed-profile phase combination. The final stage is displaced in 17 of the 18 load–phase combinations with direct baseline support, but the magnitude tracks the speed programme. Pooling loads, its median residual PC1 shift ranges from -5.0 and -6.9 at the two lowest-speed phases to -22.1 at the highest-speed phase, a factor of roughly four. This is a stage-by-speed interaction, which the additive first-order model of equation (6.4) cannot remove by construction and which the pooled R^2_{group} cannot detect because it estimates a main effect. Three of the four stages have essentially flat phase profiles on PC1, so pooling dilutes a large stage-specific dependence into the low component-level value of 0.021. The matched load–phase comparison is the only analysis in which this becomes visible.

Several robustness properties survive both qualifications. The displacement is present at every speed-profile phase and is not an artefact of unsupported load–phase combinations, since all 18 compared load–phase combinations have direct baseline support. It is not produced by baseline median imputation, as none of the audited extreme windows lacks raw mid-band descriptors. It is not specific to the feature vector, with stage association rising to 0.549 and 0.546 for the two sensitivity vectors, nor to the retained variance threshold, where it is 0.493, 0.468 and 0.459 at 80, 90 and 95 per cent. The 19 repeat windows of the first contamination sample, projected through the frozen pipeline without refitting, fall with the baseline and the first sample and an order of magnitude away from the final stage. The complete-case analysis preserves the direction of the effect, with stage association at 0.385, but it also reinforces the first qualification. Residual phase association rises to 0.253 and speed association to 0.129 when the five imputed windows are dropped, so the apparent completeness of the operational reduction depends on a small number of sparse baseline windows at the extreme speed-profile phase.

Several readings are not supported. There is no monotonic dose response, since the second contamination sample is displaced most at the lower-speed phases and almost not at all at the highest-speed phase, which is precisely where the third sample is largest. The two stages therefore respond to speed in opposite directions, which is easier to reconcile with stage-specific changes in operational sensitivity than with a single severity axis. Sparse load–phase combinations remain unstable, and at phase 5 and load 1 all three stages shift positive on the basis of two windows per stage. Increased stage association is also not improved separation, as the stage silhouette in the retained PCA score space falls from 0.095 to 0.047 under residualisation and the stage distributions continue to overlap. Baseline and the first contamination sample are not resolved from one another. The threshold-like appearance of the result should therefore not be read as a contamination threshold. A response that appears only above some severity and a response present throughout but detectable only where the operating condition amplifies it are equally compatible with the evidence, and the present design cannot separate them because contamination stage is confounded with elapsed running and acquisition order.

Residual PC1 is best described as a statistical feature contrast. Its largest loadings oppose low-band energy, counts and RMS against mid-band amplitude and energy, consistent with raw feature medians in which the final stage shows lower low-band event magnitude and higher mid band activity than the earlier stages. This is not evidence for a friction, wear, debris or lubrication mechanism. Component signs are arbitrary, the direction is a covariance artefact of this feature set, and the same contrast could arise from a change in source population, transmission path, sensor coupling or effective detection threshold at the two bands. Attribution to an AE source mechanism would require waveform-level evidence and independent condition labels, neither of which this experiment provides.

The remaining structure is therefore condition-associated only in the limited sense that it is organised by stage once measured operational main effects are reduced, that it is reproducible under frozen preprocessing, and that it is not explained by imputation, feature selection, retained dimensionality or unsupported load–phase combinations. It is not validated damage separation, not a severity scale, and not a demonstration that operational influence has been removed.

7.3.2. Interpretive Limits of PCA

PCA identifies dominant variance directions, not diagnostic classes. The reported R^2_{group} values describe organisation in the retained score space and are not detection statistics. PCA also provides an explicit projection, so new observations can be mapped into an existing score space without refitting. The resulting representation remains an analysis tool rather than a diagnostic classifier or causal model.

The evidence supports a bounded answer to the third sub-question. Measured operational structure can be identified and its first-order predictable component can be reduced. Stage-associated structure remains after that reduction within TestBlock 3, alongside operational structure that the reduction displaced rather than removed. The thesis does not show discrimination of mechanical damage, because the held-out BTF correction and the condition-varying residual comparison occur in different datasets.

PCA also provides an explicit projection, so new observations can be mapped into an existing score space without refitting. This matters for the procedure of section 7.4.2, which depends on passing later campaigns through a fixed chain: neighbour-embedding methods such as t-SNE have no out-of-sample extension, so adding observations requires the embedding to be recomputed and successive campaigns cannot be compared in the same map. A linear projection is therefore the more practical choice here, not necessarily the more sensitive one.

Whether a nonlinear representation would organise the residual feature space more usefully cannot be assessed from these data. Such methods have been applied to bearing AE features, but comparing representations requires a criterion for what counts as better separation, and that criterion needs independent condition labels at the level of the analysis window. This thesis has stage labels at run level only, so any comparison would reduce to a visual judgement. The question is therefore left to work with a labelled progression, as set out in section 7.4.5.

7.4. Implications for Offshore AE Monitoring

7.4.1. Monitoring Strategy

Raw AE trends should be interpreted relative to operating state and measurement configuration. Load, speed, direction and mechanical exposure should be recorded where relevant. Temperature and a torque or friction proxy would improve interpretation of behaviour that depends on lubrication. Thresholds, gain, screening settings and coupling information should also be logged, because they change the feature population.

No single feature is sufficient. Amplitude and waveform RMS are the strongest correction candidates. Activity measures remain useful but depend strongly on exposure and threshold. Waveform shape descriptors depend more on the sensor. Spectral descriptors provide reproducible complementary information. A multivariate representation is therefore more defensible than a single universal indicator.

Reference models should be specific to the sensor and to the operating domain. They should be fitted over the range actually observed, monitored for drift and not extrapolated. Corrected residuals are suitable quantities for trending, but this thesis does not establish diagnostic thresholds for them.

7.4.2. A Baseline Referenced Monitoring Procedure

The TestBlock results describe an analysis rather than a monitoring method. Setting out how one would become the other shows what the approach could provide and what it cannot.

Three observations support the argument. First, TestBlock 2 shows that an undamaged bearing produces substantial multivariate organisation under normal operating changes, with speed-profile phase reaching $R^2_{\text{group}} = 0.360$. Novelty detection applied to raw AE features would therefore respond to the operating programme before it responded to condition. Second, baseline-fitted residualisation reduces the pooled measure of that organisation in TestBlock 3, where speed and phase association fall from 0.334 and 0.353 to 0.060 and 0.073, although section 7.3.1 shows that part of the phase structure is displaced into low variance directions rather than removed. Third, stage-associated structure survives the reduction and comes to dominate the leading component, rising from $R^2_{\text{group}} = 0.269$ to 0.468.

The result is therefore conditional. Within one bearing, one rig and one acquisition configuration, the

predictable operational component of the feature space can be reduced, and stage-associated structure remains after that reduction.

A monitoring procedure follows in three stages. A baseline campaign is performed once on a bearing without known damage. The machine is driven through a fixed operating programme while AE and the corresponding operating variables are recorded. The window table, the per-feature operational model, the residual scaler and the principal component decomposition are then estimated and fixed. Later campaigns repeat the same operating programme and pass the new windows through the fixed chain without refitting. New observations are compared with the baseline distribution at matched load–phase combinations rather than pooled across the programme.

The comparison must use matched load–phase combinations, because the operational model is supported only where the baseline contains the corresponding load and speed-profile phase. The observed displacement also varies by a factor of roughly four across the speed programme. A departure assessed against a pooled reference would therefore depend on which load–phase combinations a campaign happened to sample. Arbitrary service data cannot be used. The monitoring unit is therefore a scheduled sweep rather than continuous acquisition, which is realistic for a slow moving bearing that cannot be opened for inspection.

The quantity reported to the operator should be scalar rather than graphical. A displacement of the campaign from the baseline centre at a matched load–phase combination, expressed in units of baseline spread, can be trended across campaigns. The PCA score space remains an analyst aid.

7.4.3. Requirements and Failure Modes

Three requirements follow from failures observed in this work rather than from general principle.

Baseline coverage must span every load–phase combination that is to be monitored later. The TestBlock 3 baseline contains no windows at load step 3 for speed-profile phases 4 and 5, so predictions there rest on the additive model alone, and the completeness of the residual reduction is sensitive to that sparse support. Thin coverage at the limits of the operating domain is the main failure mode of the procedure. Baseline coverage should therefore be strongest where the operating programme is most extreme.

Acquisition settings must stay fixed across the baseline and all later campaigns. In TestBlock 2 the detection threshold was raised from $50\,\mathrm{dB}$ to $56\,\mathrm{dB}$ and then $62\,\mathrm{dB}$ across the load sequence. Truncation alone then produced load correlations of $+0.93$ for median amplitude and $+1.00$ for RMS, and the load axis had to be withdrawn. The same retuning on a monitored asset would produce an apparent departure that could not be separated from a condition change.

The reference model is specific to the asset and to the measurement configuration, so every monitored bearing needs its own baseline campaign. This limits the practicality of the approach, because a reference cannot be supplied with the instrumentation or reused from a similar installation.

7.4.4. What the Procedure Does Not Establish

A departure from the baseline is not an attribution. The procedure indicates that the AE representation of a bearing has moved further from its own reference than the measured operating state explains. Identifying the cause remains a matter for inspection, lubricant sampling or complementary measurement.

The displacement is also not a classifier. Stage association rises after residualisation, but the stage distributions continue to overlap. No decision threshold is estimated in this work, and none should be inferred from the reported effect sizes.

The evidence supports a late stage response only. The baseline and the first contamination sample are not resolved from each other, and the second is displaced only slightly. Only the final stage is displaced consistently across the matched load–phase combinations. Whether the earlier stages produced no measurable mechanical change, or whether the representation is insensitive to changes that did occur, cannot be determined here, and no independent severity labels are available. Condition monitoring is normally justified by early warning, and early warning is not shown.

The remaining limitations of the underlying evidence are treated in section 7.6.

7.4.5. Test and Monitoring Design

Future campaigns should measure angular position directly, log calibrated load continuously and keep acquisition thresholds fixed. Test order should be balanced or randomised, and direction order alternated, so that operating effects can be separated from elapsed time. Repeated visits to identical states should be included across recordings, remounts and days.

Closing the gaps identified above requires four additions. Held-out intact data are needed to establish a false alarm rate. Several intact bearings are needed to quantify how much the reference varies between nominally identical assets. Independent condition labels are needed to interpret the displacement. A progression of known severity is needed to establish sensitivity.

The most important design improvement is to collect healthy and naturally degrading data on the same bearing with unchanged instrumentation. This would allow operational correction and preservation of condition information to be validated within one experiment.

7.5. Methodological Reflection

7.5.1. Workflow and Characterisation Choices

The thesis applies one workflow asymmetrically across heterogeneous datasets. Phase 4 provides the clearest load axis, TestBlock 2 the repeatability evidence, TestBlock 3 the stability evidence across condition states, and BTF the held-out correction test. Each campaign therefore contributes only where its design is informative. The cost is that numerical results cannot be transferred directly between rigs.

Spearman correlation was useful as a first screen, because the datasets contain few operating levels and no assumed response form. Its limitations are equally clear. BTF pressure behaviour that changes direction across the tested range can produce a weak rank correlation despite strong dependence, and pooled means can hide sign reversals between repeats. Correlation therefore functions as an association screen rather than as evidence of predictability.

7.5.2. Statistical Unit and Feature Construction

The statistical unit was chosen to reduce repeated sampling of the same operating state. Pressure–speed combinations were used at BTF and four-cycle mechanical windows in TestBlock 3. Dependence between neighbouring observations remains, so formal significance at population level is not claimed. Window length and feature vector construction remain modelling choices. Sensitivity analyses show that the main direction of the PCA result is not tied to one exact feature vector, although its numerical organisation depends on those choices.

7.6. Threats to Validity

The main construct limitation is that contamination is not mechanical damage and the BTF bearing is only a reference without known damage. Stage labels originate at run level rather than from independent condition measurements for each analysis window. Hydraulic pressure is only a proxy for the internal bearing load.

Internal validity is limited by three known confounds. Threshold is confounded with load in TestBlock 2, contamination stage with acquisition order in TestBlock 3, and both pressure and direction with test order at BTF. Lubrication state, temperature and warm-up are not measured consistently. These variables can alter the operational relationships being estimated.

Measurement uncertainty is most important at BTF and in the censored linear bearing features. BTF speed windows rely on operator logs at minute resolution, and the pressure traces are uncalibrated. Amplitude clipping and duration censoring affect parts of the linear bearing population. High-frequency populations are also small and dominated by single channels.

Statistical evidence is descriptive. Several correlations rest on only four to seven operating points, observations are dependent in time, and many feature and band combinations are screened without correction for multiplicity. The $|r_s| \geq 0.7$ criterion is therefore a trend screen rather than a significance

threshold.

Model validity is limited by sparse coverage of the operating space and by low-order mean models. Nonlinear behaviour, interactions, variance that depends on operating state, and operational relationships that depend on condition can all remain after correction. No extrapolation beyond the fitted operating domain is supported.

External validity is correspondingly limited. The linear bearing rigs and BTF do not reproduce offshore temperature, corrosion, water ingress, multi-axis loading, sea state excitation or long-term lubrication history. Numerical coefficients, thresholds and feature rankings should not be transferred directly. What transfers is the workflow. Operating state is measured, correction eligibility is tested, models are restricted to supported domains, and residuals are interpreted cautiously.

7.7. Synthesis and Boundaries of the Thesis Findings

Three findings define the thesis. First, operational sensitivity is not a fixed property of an AE feature. It depends on feature definition, frequency band, sensor path, threshold, exposure and bearing condition. Second, strong association does not imply correctability, since stability, repeatability, range coverage and predictive adequacy reduce the correctable set substantially. Third, baseline-fitted residualisation can change the dominant organisation of the multivariate feature space from operating state towards stage-associated structure, although component-level diagnostics show that operational structure is displaced rather than removed.

The evidence remains bounded. The held-out correction and the condition-varying residual comparison come from different datasets. No dataset contains a progression from healthy to damaged on the same rig, and lubrication state is not measured directly. The thesis therefore supports interpretation and correction that account for operating state as necessary components of AE monitoring. It does not support a validated damage classifier.

8

Conclusions and Recommendations

This thesis asked how the influence of operating state on the acoustic emission (AE) features of low-speed offshore rolling bearings can be characterised and reduced, so that the remaining multivariate structure can be interpreted in relation to bearing condition. One workflow was applied to four datasets: a degraded linear bearing, an intact baseline, a contamination progression and a full-scale swivel bearing without known damage. Operating state proved to be a first-order source of AE variability that can be partly modelled and reduced. This holds only for a small set of features and over a bounded operating range, and correction does not remove all operational structure.

8.1. Answers to the Research Sub-Questions

Sub-question 1 *How do measured operating parameters influence AE features in low-speed rolling bearings, and how is that influence distributed across the feature set?*

The influence is strong and unevenly distributed. It depends on the feature definition, frequency band, sensor path and bearing condition, and no single feature or band responds consistently across the datasets. Load acts most clearly on amplitude-related quantities and activity, with a mean Spearman rank correlation of $\bar{r}_s = +0.92$ for mid-band waveform RMS in the degraded linear bearing. At the swivel bearing, hit-rate rises with pressure up to about $15\,\mathrm{barg}$ and falls at higher pressure. A weak correlation over the full pressure range can therefore reflect a response that changes direction rather than insensitivity. Speed can move event rate and event magnitude in opposite directions, and whether hit-rate rises or falls with speed depends on whether it is expressed per minute or per rotation.

Contamination changes the operational relationships themselves. The low-band load coupling of acquisition RMS falls from -0.93 in the clean baseline to -0.07 at the highest contamination stage, while median amplitude remains more stable. Feature definition matters as much as the operating variable: acquisition RMS and waveform RMS agree in the direction of their load response only in the high band. Every relationship reported here depends on the detection threshold, the retained hit population and the measurement configuration.

Sub-question 2 *Which relationships between AE features and operating parameters are stable and predictable enough to support operational correction, and over what operating range can such a correction be estimated and validated?*

Only a small subset of the investigated relationships meets the complete set of criteria. Each criterion beyond association rejected at least one strongly associated feature. Rise time failed repeatability. Acquisition RMS, energy and counts failed stability across contamination stages. Crest factor and kurtosis failed held-out prediction. Median amplitude and waveform RMS were the only features to satisfy association, stability, repeatability, range coverage and predictive adequacy together. Under leave-one-pressure-out validation at the swivel bearing, they reduced the residual speed association from about -0.8 to near zero and reduced feature variance by roughly 20 to 60 per cent, depending on the channel. Their correction is supported over 0 to $15\,\mathrm{barg}$ and does not extrapolate beyond this

range. Operational correction is therefore feature-specific, sensor-specific and range-specific rather than a general transformation of the data.

Sub-question 3 *After predictable operational effects have been reduced, to what extent does condition-associated multivariate structure remain distinguishable from residual operating-state variability?*

To a limited extent. Before correction, the contamination feature space is organised mainly by operating state rather than by contamination stage. The intact reference shows a comparable organisation without any change in condition, so raw multivariate structure cannot be interpreted as condition-related. After the predictable operational component is removed, contamination stage becomes the dominant organising factor. The fraction of retained PCA score variation associated with stage, R^2_{group}, rises from 0.269 to 0.468, while the fraction associated with speed and speed-profile phase falls below 0.08. This pattern is retained across the alternative feature vectors and robustness checks performed.

Separability is limited by operational structure that the correction does not model. Phase structure persists in lower-variance components, and the displacement of the most contaminated stage itself depends on speed, an interaction that an additive model cannot remove. The stage distributions also continue to overlap, the baseline and lowest contamination stage are not resolved, and the response is not monotonic with contamination level. Because contamination stage is confounded with acquisition order, the remaining structure is stage-associated rather than attributable to contamination alone. Condition-associated structure is therefore distinct enough to change the dominant organisation of the feature space, but not enough to support a severity scale or decision threshold.

8.2. Answer to the Main Research Question

The main research question was

How can the influence of operating state on acoustic emission features be characterised and reduced so that the remaining multivariate structure can be interpreted in relation to bearing condition in low-speed offshore rolling bearings?

The influence of operating state can be characterised by screening each feature against the measured operating variables and requiring stability, repeatability, range coverage and held-out predictive adequacy in addition to association. Its predictable component can be reduced by low-order regression fitted on a reference condition, but only for features that meet every criterion and only within the covered operating range. Residualisation removes the modelled mean response rather than all operational structure. Within these limits, the quantity of interest for condition interpretation becomes the departure from a bearing's own operationally corrected reference rather than the raw feature level.

This answer has three important boundaries. The validated correction was demonstrated on a bearing without known damage, whereas the condition-varying comparison relied on a baseline-fitted residualisation that could not be validated on held-out operating states. Neither half of the workflow therefore validates the other. No dataset used here pairs a healthy and a mechanically damaged state on the same rig, and contamination stage is confounded with acquisition order. Lubrication state was not measured directly and was shown to change the relationships that a correction assumes to remain stable. The thesis therefore does not establish a validated diagnostic classifier. Within these boundaries, it establishes an operating-state-aware framework for interpreting AE measurements relative to a bearing's own reference behaviour, and supports this as a more defensible basis for condition monitoring than raw feature trending.

8.3. Contributions

The main contributions of this work are

- **A correction eligibility framework** that applies association, stability, repeatability, range coverage and predictive adequacy in sequence. A criterion is recorded as not testable when the experimental design does not support its evaluation. This allows datasets with different designs to contribute without assuming unsupported evidence.
- **Evidence that association strength alone is a poor guide to correctability**. Several strongly

associated features failed later eligibility criteria. The analysis also showed that acquisition RMS and waveform RMS are distinct measurements whose operational responses can lead to different conclusions.

- **Component-level and matched operating-state diagnostics for residual PCA**. These show that a pooled reduction in operational association does not necessarily mean that operational structure has been removed from the complete feature space.
- **A modular Python workflow** for AE and operational data, covering cleaning, synchronisation, feature extraction, operational characterisation, correction and PCA. The workflow was applied across four heterogeneous datasets.

The framework and its criteria are proposed as transferable. The fitted coefficients and feature rankings are not.

8.4. Recommendations for Monitoring and Test Design

1. **Log operating state together with AE.** Record angular position or displacement, direction and a calibrated load or pressure signal where relevant. Temperature should also be recorded, and a torque or friction proxy is valuable where available.
2. **Keep acquisition settings fixed and record them.** Detection threshold, gain, screening rules and coupling conditions affect the retained AE population. Threshold retuning during one campaign in this work produced apparent load correlations of up to $+1.00$ through truncation alone.
3. **Build a reference for each asset, sensor and measurement configuration.** A correction did not transfer between nominally equivalent sensors on the same bearing. A baseline campaign should populate every operating state that will later be monitored. Later measurements should then be compared with the baseline at matched operating states rather than pooled across the operating programme, since pooled association measures concealed an operating-dependent stage response in this work. Observations outside the supported range should be treated as out of domain rather than scored against the reference, and the reference itself should be monitored for drift.
4. **Separate the features that are corrected from those that are trended.** Median amplitude and waveform RMS are predictable from operating state, which makes them candidates for correction rather than health indicators in their own right. Their sensitivity to damage has not been demonstrated. Activity measures carry information but were the least repeatable quantities measured. Spectral descriptors were reproducible and weakly coupled to operating state, which makes them candidates for carrying condition information.
5. **Design experiments so that factors are not confounded with order.** In this work, contamination stage could not be separated from acquisition order. Test order should be balanced or randomised, direction order should alternate between speed levels, and load and speed should be varied within a recording where the rig allows. Identical operating states should be revisited across recordings, remounts and days.

8.5. Recommendations for Further Research

1. **Apply the workflow to available run-to-failure datasets.** Datasets containing natural degradation on the same bearing, with concurrent operational records that are available or can be reconstructed, would test whether departures from an early-life reference develop consistently as damage progresses. This would address the absence of a paired healthy and damaged state in the present evidence.
2. **Model interactions between condition and operating state.** The stage displacement observed here changes with speed. An additive main-effect model cannot represent this behaviour. Future correction models should allow operational sensitivity itself to change with condition.
3. **Measure or estimate lubrication state.** Lubricant condition altered relationships that the correction assumes to remain stable. Measuring this state would help distinguish a changing lubricant condition from an outdated operational model.

4. **Extend correction beyond the low-order model.** The swivel-bearing pressure response changes direction within the measured range, and the contamination residualisation lacked baseline observations at some combinations of load step and speed-profile phase. Denser operating coverage would allow nonlinear or regime-dependent models to be evaluated without extrapolation, and would show how much of the phase structure remaining in the lower-variance components a better-specified model can remove.
5. **Establish the quantities required for monitoring decisions.** Held-out intact data are needed to estimate false-alarm behaviour. Several nominally identical assets are needed to quantify between-bearing variation, while independent condition information is required to interpret departures from the reference.
6. **Test the workflow on independent offshore data.** The sequence of operational characterisation, eligibility testing, correction and residual analysis should be evaluated outside the laboratory to determine whether the same interpretation remains useful under service conditions.
7. **Compare the approach with waveform similarity analysis.** Waveform similarity analysis assumes a stable transmission path, whereas the approach taken here assumes that the governing operating variables are measured. Both should be applied to the same reference and condition-changing data to establish whether they provide complementary information.

References

[1] Bart Scheeren, Miroslaw Lech Kaminski, and Lotfollah Pahlavan. “Acoustic emission monitoring of naturally developed damage in large-scale low-speed roller bearings”. In: *Structural Health Monitoring* 23.1 (2024), pp. 360–382. DOI: 10.1177/14759217231164912.

[2] Henrique Dias Machado de Azevedo, Alex Maurício Araújo, and Nadège Bouchonneau. “A review of wind turbine bearing condition monitoring: State of the art and challenges”. In: *Renewable and Sustainable Energy Reviews* 56 (2016), pp. 368–379. DOI: 10.1016/j.rser.2015.11.032.

[3] Wahyu Caesarendra et al. “Acoustic emission-based condition monitoring methods: Review and application for low speed slew bearing”. In: *Mechanical Systems and Signal Processing* 72–73 (2016), pp. 134–159. DOI: 10.1016/j.ymssp.2015.10.020.

[4] Idriss El-Thalji and Erkki Jantunen. “A descriptive model of wear evolution in rolling bearings”. In: *Engineering Failure Analysis* 45 (2014), pp. 204–224. DOI: 10.1016/j.engfailanal.2014.06.004.

[5] A. Cockerill et al. “Determination of rolling element bearing condition via acoustic emission”. In: *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology* 230.11 (2016), pp. 1377–1388. DOI: 10.1177/1350650116638612.

[6] R. Fuentes et al. “Detection of sub-surface damage in wind turbine bearings using acoustic emissions and probabilistic modelling”. In: *Renewable Energy* 147 (2020), pp. 776–797. DOI: 10.1016/j.renene.2019.08.019.

[7] J. D. Smith. “Vibration monitoring of bearings at low speeds”. In: *Tribology International* 15 (1982), pp. 139–144.

[8] N. Jamaludin, D. Mba, and R. H. Bannister. “Condition monitoring of slow-speed rolling element bearings using stress waves”. In: *Proceedings of the Institution of Mechanical Engineers, Part E: Journal of Process Mechanical Engineering* 215.4 (2001), pp. 245–271.

[9] Takeo Yoshioka and Takashi Fujiwara. “A new acoustic emission source locating system for the study of rolling contact fatigue”. In: *Wear* 81 (1982). Short Communication, pp. 183–186.

[10] Alan Hase, Hiroshi Mishina, and Masaki Wada. “Correlation between features of acoustic emission signals and mechanical wear mechanisms”. In: *Wear* 292–293 (2012), pp. 144–150. DOI: 10.1016/j.wear.2012.05.019.

[11] N. Tandon and B. C. Nakra. “Comparison of vibration and acoustic measurement techniques for the condition monitoring of rolling element bearings”. In: *Tribology International* 25.3 (1992), pp. 205–212.

[12] M. Elforjani and D. Mba. “Monitoring the Onset and Propagation of Natural Degradation Process in a Slow Speed Rolling Element Bearing With Acoustic Emission”. In: *Journal of Vibration and Acoustics* 130.4 (2008), p. 041013. DOI: 10.1115/1.2948413.

[13] Hoon Sohn. “Effects of environmental and operational variability on structural health monitoring”. In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 365 (2007), pp. 539–560. DOI: 10.1098/rsta.2006.1935.

[14] Eloi Figueiredo and James Brownjohn. “Three decades of statistical pattern recognition paradigm for SHM of bridges”. In: *Structural Health Monitoring* 21.6 (2022), pp. 3018–3054. DOI: 10.1177/14759217221075241.

[15] Bart Scheeren. “On Acoustic Emission Condition Monitoring of Highly-loaded Low-speed Roller Bearings”. PhD thesis. Delft University of Technology, 2023. DOI: 10.4233/uuid:e3c4bcaa-e7fa-499c-85b0-8cb5d4f473d8.

[16] Zepeng Liu et al. “Acoustic Emission Analysis for Wind Turbine Blade Bearing Fault Detection Under Time-Varying Low-Speed and Heavy Blade Load Conditions”. In: *IEEE Transactions on Industry Applications* (2021). DOI: 10.1109/TIA.2021.3058557.

[17] Hong Guo, Lindsay B. Jack, and Asoke K. Nandi. “Feature Generation Using Genetic Programming with Application to Fault Classification”. In: *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics* 35.1 (2005), pp. 89–99. DOI: 10.1109/TSMCB.2004.841426.

[18] Bart Scheeren, Nitish P. Thakoerdajal, and Lotfollah Pahlavan. “Influence of Grease Contamination on Acoustic Emission Monitoring of Low-speed Roller Bearings”. In: *e-Journal of Nondestructive Testing* (2024), pp. 1–8. DOI: 10.58286/29672.

[19] Ahmed Abdelbary and Li Chang. “Failure Analysis of Tribo-Systems”. In: *Principles of Engineering Tribology: Fundamentals and Applications*. Academic Press, 2023, pp. 521–562. ISBN: 978-0-323-99115-5. DOI: 10.1016/B978-0-323-99115-5.00015-3.

[20] SKF. *Bearing Damage and Failure Analysis*. Technical handbook. Gothenburg, Sweden: SKF Group, n.d.

[21] Farshid Sadeghi et al. “A Review of Rolling Contact Fatigue”. In: *Journal of Tribology* 131.4 (2009), p. 041403. DOI: 10.1115/1.3209132.

[22] N. Symonds. *Rolling Contact Fatigue – Review and Case Study*. Tech. rep. RTO-AVT-109. NATO RTO, 2004.

[23] Pawel Rycerz, Andrew Olver, and Amir Kadiric. “Propagation of surface initiated rolling contact fatigue cracks in bearing steel”. In: *International Journal of Fatigue* 97 (2017), pp. 29–38. DOI: 10.1016/j.ijfatigue.2016.12.004.

[24] J. Halme and P. Andersson. “Rolling contact fatigue and wear fundamentals for rolling bearing diagnostics – state of the art”. In: *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology* 224.4 (2009), pp. 377–395. DOI: 10.1243/13506501JET656.

[25] A. Choudhury and N. Tandon. “Application of acoustic emission technique for the detection of defects in rolling element bearings”. In: *Tribology International* 33 (2000), pp. 39–45.

[26] Zhaowen Xu et al. “Review on Structural Health Monitoring of Pipeline: From Perspective on Acoustic Based Crack Detection”. In: *Measurement* 263 (2026), p. 120075. DOI: 10.1016/j.measurement.2025.120075.

[27] Arturo Silva-Campillo, M. A. Herreros-Sierra, and Francisco Pérez-Arribas. “Acoustics as a Structural Health Monitoring Tool in Naval and Offshore Structures: A Comprehensive Review”. In: *Applied Sciences* 16.3 (2026), p. 1477. DOI: 10.3390/app16031477.

[28] Ali Mardanshahi et al. “Sensing Techniques for Structural Health Monitoring: A State-of-the-Art Review on Performance Criteria and New-Generation Technologies”. In: *Sensors* 25.5 (2025), p. 1424. DOI: 10.3390/s25051424.

[29] J. Couturier and D. Mba. “Operational Bearing Parameters and Acoustic Emission Generation”. In: *Journal of Vibration and Acoustics* 130.2 (2008), p. 024502. DOI: 10.1115/1.2776339.

[30] Bart Scheeren and Pooria Pahlavan. *Report Phase 1: HiTeAM – High-Tech Acoustic Monitoring*. Tech. rep. Version 2.0. Version 2.0, 17 June 2019. HiTeAM Joint Industry Programme, June 2019.

[31] M. Elforjani and D. Mba. “Accelerated natural fault diagnosis in slow speed bearings with Acoustic Emission”. In: *Engineering Fracture Mechanics* 77 (2010), pp. 112–127. DOI: 10.1016/j.engfracmech.2009.09.016.

[32] Jiaojiao Ma et al. “Condition monitoring of journal bearings based on acoustic emissions: A state-of-the-art review”. In: *Friction* 14.1 (2026), p. 9441080. DOI: 10.26599/FRICT.2025.9441080.

[33] Bart Scheeren et al. *Report Phase 2: HiTeAM – High-Tech Acoustic Monitoring*. Tech. rep. Version 2.0. Version 2.0, 13 December 2021. HiTeAM Joint Industry Programme, Dec. 2021.

[34] Bart Scheeren, Miroslaw Lech Kaminski, and Lotfollah Pahlavan. “Evaluation of Ultrasonic Stress Wave Transmission in Cylindrical Roller Bearings for Acoustic Emission Condition Monitoring”. In: *Sensors* 22.4 (2022), p. 1500. DOI: 10.3390/s22041500.

[35] M. Elforjani and D. Mba. “Detecting natural crack initiation and growth in slow speed shafts with the Acoustic Emission technology”. In: *Engineering Failure Analysis* 16.7 (2009), pp. 2121–2129. DOI: 10.1016/j.engfailanal.2009.02.005.

[36] Mohammadhadi Shateri et al. “On Acoustic Emission for Damage Detection and Failure Prediction in Fiber Reinforced Polymer Rods Using Pattern Recognition Analysis”. In: *Smart Materials and Structures* 26.6 (2017), p. 065023. DOI: 10.1088/1361-665X/aa6e43.

[37] A. M. Al-Ghamdi et al. “Estimation of bearing defect size with acoustic emission”. In: *Insight* 46.12 (2004), pp. 758–761.

[38] Saad Al-Dossary, R. I. Raja Hamzah, and D. Mba. “Observations of Changes in Acoustic Emission Waveform for Varying Seeded Defect Sizes in a Rolling Element Bearing”. In: *Applied Acoustics* 70.1 (2009), pp. 58–81. DOI: 10.1016/j.apacoust.2008.01.005.

[39] Juha Miettinen and Pentti Pataniitty. “Acoustic Emission in Monitoring Extremely Slowly Rotating Rolling Bearing”. In: *Proceedings of COMADEM ’99*. Oxford: Coxmoor Publishing Company, 1999, pp. 289–297. ISBN: 1-901892-13-1.

[40] Yi Guo et al. *Acoustic Emission Measurement of a Wind Turbine Main Bearing*. Tech. rep. NREL/TP-5000-83370. Golden, CO: National Renewable Energy Laboratory, Oct. 2022. URL: https://www.nrel.gov/docs/fy23osti/83370.pdf.

[41] Achmad Widodo et al. “Fault diagnosis of low speed bearing based on acoustic emission signal and multi-class relevance vector machine”. In: *Proceedings of ICSV15*. Daejeon, Korea, July 2008, pp. 1468–1475.

[42] Brandon Van Hecke, Jae Yoon, and David He. “Low Speed Bearing Fault Diagnosis Using Acoustic Emission Sensors”. In: *Applied Acoustics* 105 (2016), pp. 35–44. DOI: 10.1016/j.apacoust.2015.10.028.

[43] D. Mba, R. H. Bannister, and G. E. Findlay. “Condition monitoring of low-speed rotating machinery using stress waves. Part 1”. In: *Proceedings of the Institution of Mechanical Engineers, Part E: Journal of Process Mechanical Engineering* 213.3 (1999), pp. 153–170.

[44] Yongyong He, Xinming Zhang, and Michael I. Friswell. “Defect Diagnosis for Rolling Element Bearings Using Acoustic Emission”. In: *Journal of Vibration and Acoustics* 131.6 (2009), p. 061012. DOI: 10.1115/1.4000480.

[45] Daniel Cornel et al. “Condition monitoring of roller bearings using acoustic emission”. In: *Wind Energy Science* 6 (2021), pp. 367–376. DOI: 10.5194/wes-6-367-2021.

[46] Juha Miettinen and Peter Andersson. “Acoustic Emission of Rolling Bearings Lubricated with Contaminated Grease”. In: *Tribology International* 33.11 (2000), pp. 777–787. DOI: 10.1016/S0301-679X(00)00124-9.

[47] Lizzie Neumann, Philipp Wittenberg, and Jan Gertheiss. *Removal of Multivariate Environmental Influences in Structural Health Monitoring through Conditional Covariances and Supervised Learning*. 2026. arXiv: 2606.05026 [stat.ME].

[48] K. Worden, H. Sohn, and C. R. Farrar. “Novelty detection in a changing environment: regression and interpolation approaches”. In: *Journal of Sound and Vibration* 258.4 (2002), pp. 741–761. DOI: 10.1006/jsvi.2002.5148.

[49] A.-M. Yan et al. “Structural damage diagnosis under varying environmental conditions—part II: local PCA for non-linear cases”. In: *Mechanical Systems and Signal Processing* 19 (2005), pp. 865–880. DOI: 10.1016/j.ymssp.2004.12.003.

[50] Vallen Systeme GmbH. *AMSY-6 System Description*. System Description, document released 2021-11. Vallen Systeme GmbH. Wolfratshausen, Germany, Nov. 2021.

[51] Charles R. Harris et al. “Array programming with NumPy”. In: *Nature* 585.7825 (2020), pp. 357–362. DOI: 10.1038/s41586-020-2649-2.

[52] Wes McKinney. “Data Structures for Statistical Computing in Python”. In: *Proceedings of the 9th Python in Science Conference*. 2010, pp. 56–61. DOI: 10.25080/Majora-92bf1922-00a.

[53] Pauli Virtanen et al. “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python”. In: *Nature Methods* 17 (2020), pp. 261–272. DOI: 10.1038/s41592-019-0686-2.

[54] Skipper Seabold and Josef Perktold. “statsmodels: Econometric and Statistical Modeling with Python”. In: *Proceedings of the 9th Python in Science Conference*. 2010.

[55] Fabian Pedregosa et al. “Scikit-learn: Machine Learning in Python”. In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[56] Patrick Schober, Christa Boer, and Lothar A. Schwarte. “Correlation Coefficients: Appropriate Use and Interpretation”. In: *Anesthesia & Analgesia* 126.5 (2018), pp. 1763–1768. DOI: 10.1213/ANE.0000000000002864.

[57] Haldun Akoglu. “User’s guide to correlation coefficients”. In: *Turkish Journal of Emergency Medicine* 18.3 (2018), pp. 91–93. DOI: 10.1016/j.tjem.2018.08.001.

[58] B. Scheeren and L. Pahlavan. *Report Phase 4: HiTeAM – High-Tech Acoustic Monitoring*. Tech. rep. Version 1.0. HiTeAM Joint Industry Programme, Apr. 24, 2025.

[59] Ian T. Jolliffe. *Principal Component Analysis*. 2nd ed. Springer Series in Statistics. New York: Springer, 2002. DOI: 10.1007/b98835.

[60] Peter J. Rousseeuw and Mia Hubert. “Anomaly Detection by Robust Statistics”. In: *WIREs Data Mining and Knowledge Discovery* 8 (2018), e1236. DOI: 10.1002/widm.1236.

[61] Marti J. Anderson. “A New Method for Non-Parametric Multivariate Analysis of Variance”. In: *Austral Ecology* 26.1 (2001), pp. 32–46. DOI: 10.1111/j.1442-9993.2001.01070.pp.x.

A

Research Paper

The research paper starts on the next page.

Operational Normalisation of Acoustic Emission Features in Low-Speed Offshore Rolling Bearings

A. M. Koetsier[a], R. Habiyaremye[a], L. Pahlavan[a], B. Scheeren[b], N. Daane[c], and J. Jovanova[a]
[a] Department of Maritime and Transport Technology, Faculty of Mechanical Engineering, Delft University of Technology, Delft, The Netherlands
[b] HiTeAM Systems, Delft, The Netherlands
[c] Bluewater Energy Services, Hoofddorp, The Netherlands

***Abstract*—Acoustic emission (AE) features of low-speed rolling bearings respond to operating state as well as to bearing condition, so an operational change can be mistaken for a change in condition. This paper investigates how operational influence can be characterised and reduced before multivariate AE measurements are interpreted. One workflow is applied to four datasets: a degraded linear bearing, an intact reference, a grease contamination progression and a full-scale swivel bearing without known damage. A feature is corrected only when its operational relationship is stable, repeatable, supported by the sampled range and predictive on held-out data. Operational sensitivity depended strongly on feature definition, frequency band and measurement configuration, and contamination altered the operational relationships themselves. On the swivel bearing, only median amplitude and waveform RMS met all criteria. Over 0–$15\,\mathrm{barg}$, leave-one-pressure-out validation reduced their speed correlations from between -0.70 and -0.88 to between $+0.02$ and $+0.25$, and reduced their variance by 19–$61\,\%$. In the contamination dataset, baseline-fitted residualisation lowered the share of retained PCA score variation associated with speed and speed-profile phase from 0.33–0.35 to 0.06–0.07, while the stage-associated share rose from 0.269 to 0.468. Residual operational structure nevertheless remained. Operational correction is therefore feature-, sensor- and range-specific. For monitoring practice, the results support comparing a bearing with its own operationally corrected baseline at matched operating states instead of trending raw AE features. They do not establish a validated diagnostic classifier.**

***Index Terms*—Acoustic emission, condition monitoring, low-speed rolling bearings, operational variability, data normalisation, principal component analysis.**

I. Introduction

ROLLING bearings in offshore heavy-lifting cranes, mooring systems and wind energy installations operate under high loads, intermittent or oscillatory motion and very low rotational speeds, with limited inspection access [1], [2]. Fatigue, wear, lubrication distress and debris damage may develop concurrently in such bearings [3], so the measured response reflects several interacting processes.

Vibration-based diagnosis becomes unreliable below approximately $100\,\mathrm{rpm}$ [4]. Acoustic emission (AE) measures the transient elastic stress waves released by crack initiation, plastic deformation, friction and asperity interaction, and its applicability to low-speed bearings is well documented [1], [5], [6]. The same sensitivity complicates its interpretation. Speed is the most consistently reported operational influence and has been linked to the specific lubricant film thickness, whereas the load response is weaker and less consistent [7], [8]. Grease contamination raises AE activity at very low particle concentrations [9], [10], and field measurements show that several operating parameters vary together in service [11]. A change in an AE feature can therefore not be attributed to degradation until the expected response at the prevailing operating state is known. For an operator, an alarm on a raw AE trend may reflect a change in load or speed and not a change in the bearing.

This is the data normalisation problem of structural health monitoring (SHM), in which environmental and operational variability can mask or be mistaken for damage [12], [13]. When the governing variables are measured, the reference response can be modelled explicitly by regression and features interpreted relative to it [14], [15]. Principal component analysis (PCA) has been applied to AE features for low-speed bearing diagnosis [16]. PCA follows directions of maximum variance, however, and does not establish whether the recovered structure originates in operating state or in condition [13], [17].

This paper addresses that gap for low-speed offshore rolling bearings. Operational sensitivity is first characterised across two bearing geometries. A correction-eligibility procedure is then applied that requires stability, repeatability, range coverage and held-out predictive adequacy in addition to association. Finally, raw and residual feature spaces are compared using PCA. No fault classifier is developed; the aim is to determine whether operating-state-aware processing provides a more defensible basis for interpreting AE measurements than raw feature trending.

II. Experimental Datasets

Four datasets are used (Table I), where TB2 and TB3 denote TestBlocks 2 and 3 of the HiTeAM Phase 2 campaign. They differ in bearing type, scale, sensor layout and operating programme. Each is therefore processed independently and used only for the analytical role that its design supports. All were acquired with a Vallen AMSY-6 system at $5\,\mathrm{MHz}$ in three bands of 40–$100\,\mathrm{kHz}$, 95–$180\,\mathrm{kHz}$ and 180–$580\,\mathrm{kHz}$, referred to as the low, mid and high bands.

The swivel bearing experiment was conducted for this work. Axial load was applied by pressurising the internal water

Corresponding author: a.m.koetsier Manuscript pre-pared September 2026 as MSc thesis report 2026.MME.9219.

TABLE I
DATASETS AND THEIR ANALYTICAL ROLES

Dataset	Bearing and condition	Operating programme	Hits	Role and main limitation
Phase 4 [18]	Double linear bearing, end of service life, severe degradation confirmed	18 load sweeps over 7 load levels. 3 stroke speeds and 3 lubrication conditions between recordings	7.9×10^6	Load characterisation in an advanced degradation state. No revisited states
TB2 [19]	Linear bearing rig, pristine raceways and rollers, intact	4 loads (100–1305 kN), 3 stroke speeds visited ascending and descending	1.9×10^6	Speed characterisation, repeatability, intact PCA reference. Load axis withdrawn
TB3 [19]	Same rig, baseline and three Fe contamination stages (Samples 1–3: 16, 800 and 40 000 $\mathrm{mg\,kg^{-1}}$)	Same loads and speeds in every stage, threshold stable near 51 dB	5.4×10^6	Stability across condition states, raw and residual PCA. Stage confounded with order
Swivel bearing	Full-scale three-row roller bearing, no known damage, about 2.5×10^5 equivalent revolutions	6 pressures (0–25 barg, up to 1817 kN), 5 speeds (0.5–6.0 $^\circ\,\mathrm{s^{-1}}$) in both directions	185 013 (135 481 kept)	Full eligibility sequence with held-out validation. Operator-log timing, pressure confounded with test order

volume. Three sensor nodes at 120° spacing on the outer ring each carried a low-, mid- and high-band transducer. The interior was not inspected, so the bearing is treated as a run-in reference without known damage and not as demonstrably unworn.

In TB3, contamination increased sequentially without cleaning between stages, so stage is confounded with acquisition order. In TB2, the detection threshold of the dominant channels was raised by about 12 dB across the load sequence. This retuning alone produced apparent load correlations of $+0.93$ for amplitude and $+1.00$ for waveform RMS, and the TB2 load axis was therefore withdrawn.

III. METHOD

The workflow is shown in Fig. 1. Where a dataset does not support held-out validation, the correction stage is applied as an exploratory residualisation and reported as such.

A. Observations and Features

Hits are aligned with the operational records and grouped by nominal operating state. The feature set combines acquisition, waveform, spectral and activity descriptors, aggregated by the median within each analysis unit and frequency band. Acquisition RMS represents the pre-trigger background level [20] and is kept separate from waveform RMS of the stored transient. For the swivel bearing, a band-specific amplitude gate retains 73 % of the hits. Of these, 85 % are mid-band, so the mid band carries the correction analysis.

B. Operational Characterisation and Eligibility

Association is quantified by the Spearman rank correlation, because the campaigns contain few operating levels and no functional form is assumed. Coefficients are computed per repeat of the varied axis and averaged. A feature and band combination is screened as operationally associated when $|\bar{r}_s| \geq 0.7$. The number of repeats meeting the screen is reported with the mean, since a pooled mean can conceal sign reversals.

Four further criteria decide whether a relationship supports correction. *Stability* requires the direction to persist across repeats. It is summarised by the fraction ϕ of repeat coefficients that share the sign of the mean, and by their spread. *Repeatability* requires a reproducible value where a commanded state is revisited. It is quantified by the symmetric relative difference $d = 200\,|v_a - v_b|/(|v_a| + |v_b|)$ between the feature values v_a and v_b at the two visits. A median d below 10 % is classed as repeatable and above 25 % as unreliable. *Range coverage* restricts estimation to an adequately populated operating range. *Predictive adequacy* requires the fitted model to predict reference observations withheld from estimation. All thresholds are study-specific screening values and are not significance levels.

C. Correction

Following regression-based SHM normalisation [12], [14], [15], the expected value of feature x_j at the measured operating state $\mathbf{z}$ is estimated on reference data and the residual is carried forward,

$$\hat{x}_j = f_j(\mathbf{z}), \qquad r_j = x_j - \hat{x}_j. \tag{1}$$

Condition labels are never used to estimate f_j. For the swivel bearing, a linear additive pressure and speed model is compared with an interaction model per feature and channel. The correction is restricted to 0–15 barg. Minor flange leakage occurred at 20 and 25 barg, where pressure required manual adjustment, whereas no leakage was reported at 15 barg or below. The two highest-pressure recordings are therefore retained for full-range characterisation but are not used to estimate or validate the correction model. Their feature responses also depart from the trend observed over 0–15 barg. Predictive adequacy within the retained range is assessed by leave-one-pressure-out validation.

TB3 offers no held-out structure. Each feature, log-transformed except for amplitudes in decibels, is instead regressed on categorical load step and speed-profile phase using baseline windows only. The speed-profile phase is the position within the $0 \to 1 \to 2 \to 1 \to 0$ speed programme of each recording. Interaction terms are omitted because two load and phase combinations are absent from the baseline. Residuals are centred and scaled by the baseline median and interquartile range [21].

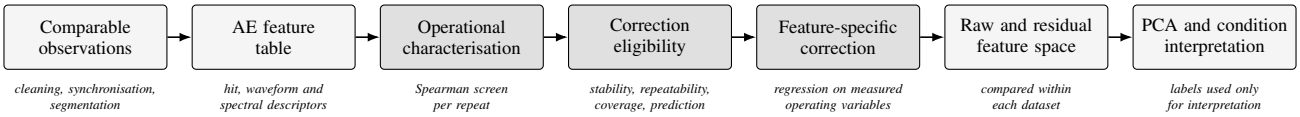


Fig. 1. Analysis workflow. Shaded stages constitute the operational normalisation. Ineligible features bypass the correction and remain descriptive.

D. Multivariate Analysis

The TB2 and TB3 observations are non-overlapping windows of four accepted mechanical cycles, which retains 92.3 % of accepted cycles. The primary analysis uses 294 TB3 windows and a 13-feature core vector: low- and mid-band hit rate, median amplitude, rise time, energy, RMS and counts, together with mid-band duration. Low-band duration is excluded because it is censored by the acquisition limit. High-band descriptors are excluded because they are undefined in the 70 windows without high-band hits. Five windows without mid-band hits are imputed from baseline medians. Two sensitivity vectors test the dependence on this selection: an extended vector that adds hit counts and percentile descriptors, and the core vector with high-band activity added.

PCA is applied to the robust-scaled raw and residual vectors, with components retained up to 90 % cumulative variance [22]. Organisation of the retained scores $\mathbf{z}_i$ by a label with groups g is summarised by the between-group share of score variation [23],

$$R^2_{\text{group}} = \frac{\sum_g n_g \|\bar{\mathbf{z}}_g - \bar{\mathbf{z}}\|^2}{\sum_i \|\mathbf{z}_i - \bar{\mathbf{z}}\|^2}, \tag{2}$$

which is used descriptively and not as a classification metric.

IV. Results

A. Operational Sensitivity Is Feature- and Band-Specific

Table II summarises the principal relationships. In Phase 4, load sensitivity is strong but no feature is load-associated in all three bands. High-band hit rate, mid-band waveform RMS and mid-band amplitude reach $+0.89$, $+0.92$ and $+0.89$ (Fig. 3a). Acquisition RMS, in contrast, is positive in the high band and negative in the low band. At constant load, a higher stroke speed raises high-band hit rate by a factor of 1.95, while high-band RMS and peak amplitude fall by factors of 2.96 and 3.39, so event rate and event magnitude respond differently to speed.

In the swivel bearing, no feature and band mean reaches the screen over the complete 0–25 barg sweep, whereas the 0–15 barg range gives hit-rate correlations of $+0.88$ to $+0.96$. Hit rate is unimodal in pressure (Fig. 2), so the weak full-range coefficient reflects a response that changes direction rather than insensitivity. Pressure changes event occurrence far more than event magnitude, as hit rate spans 13.5–1806 hits/min whereas median amplitude stays within 51.6–53.4 dB.

Speed acts on a different part of the feature set. Amplitude-related quantities fall with speed in all bands. Mid-band hit rate rises with speed when expressed per minute but falls when expressed per nominal rotation, from about 229 to 108 hits. The exposure definition is therefore part of the feature definition. The sign of the amplitude response to speed also differs between rigs, being positive in TB2 and TB3 (Fig. 3b) and negative in the swivel bearing.

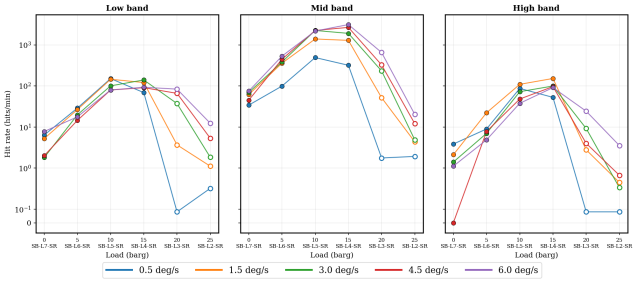


Fig. 2. Swivel bearing hit rate versus pressure by frequency band. Filled markers denote the 0–15 barg range that supports correction, open markers the excluded 20 and 25 barg recordings.

The retained hit population conditions every relationship. In TB2, a signal-to-noise filter that retains 0.45 % of low-band hits reverses the speed correlation of waveform RMS from $+0.88$ to -1.00.

B. Contamination Alters the Operational Relationships

Table III decomposes the TB3 low-band load correlations by stage. Median amplitude and both derived ratios retain their load relationship in every stage (Fig. 3c). Acquisition RMS changes from -0.93 in the baseline to -0.07 at the highest contamination level (Fig. 3d), and energy, counts and duration weaken at that level.

At matched operating states the event population changes as well. Between the baseline and Sample 3, hit rate approximately doubles at every load. Mean duration falls from 13 407 µs to 6921 µs and counts from 491 to 233, while amplitude changes only from 57.5 dB to 58.9 dB. Median mid-band activity rises from 54 to 641 hits per window. The later stages therefore contain more and shorter events at similar amplitude.

C. Association Does Not Determine Correctability

Table IV lists the eligibility outcomes. In the swivel bearing, only median amplitude and waveform RMS satisfy the complete sequence. Their raw speed correlations of about -0.70 to -0.88 across the three mid-band channels fall to $+0.02$ to $+0.25$ under leave-one-pressure-out validation, with variance reductions of 19–61 %.

Each criterion beyond association rejected at least one strongly associated feature. Hit rate gives the strongest pressure association in the swivel bearing but differs by 46–81 % between anticlockwise and clockwise visits to the same commanded state, whereas median amplitude differs by at most 0.6 %. Timing uncertainty in the operator log explains part of this difference but not the large within-pass variation. Crest factor and kurtosis are associated and repeatable, yet

TABLE II
SELECTED OPERATIONAL SENSITIVITIES. MEAN SPEARMAN CORRELATION WITH THE NUMBER OF REPEATS SATISFYING $|r_s| \geq 0.7$ IN PARENTHESES

Linear bearing datasets				Swivel bearing			
Dataset	Axis	Feature (band)	$\bar{r}_s$	Sweep	Axis	Feature (band)	$\bar{r}_s$
Phase 4	Load	Hit rate (high)	+0.89 (16/18)	Full	Pressure	Hit rate (mid)	−0.28 (0/5)
		Hit rate (low)	+0.45 (8/18)			Kurtosis (mid)	−0.81 (4/5)
		Waveform RMS (mid)	+0.92 (17/18)	Reduced	Pressure	Hit rate (low / mid)	+0.92 / +0.88
		Amplitude (mid)	+0.89 (17/18)			Hit rate (high)	+0.96 (5/5)
		Acquisition RMS (low)	−0.59 (11/18)			Counts (mid)	+0.86 (5/5)
		Acquisition RMS (high)	+0.84 (14/18)			Amplitude (mid)	+0.04 (3/5)
TB2	Speed	Amplitude, energy, rise time (low)	+1.00 (4/4)	Reduced	Speed	Amplitude (low)	−0.93 (4/4)
		Waveform RMS (low)	+0.88 (3/4)			Amplitude (mid)	−0.82 (3/4)
TB3	Load	Amplitude (low)	+0.97 (12/12)			Waveform RMS (mid)	−0.95 (4/4)
		Waveform RMS (low)	+0.00 (5/12)			Energy (mid)	−0.95 (4/4)
TB3	Speed	Amplitude (low)	+1.00 (16/16)			Hit rate per minute (mid)	+0.82 (3/4)
		Acquisition RMS (low)	−1.00 (16/16)			Hit rate per rotation (low)	−0.87 (4/4)
		Hit rate (mid)	+0.91 (15/16)			Hit rate per rotation (high)	−0.92 (4/4)

Full sweep 0–25 barg, reduced sweep 0–15 barg. Swivel bearing values use the amplitude-screened population.

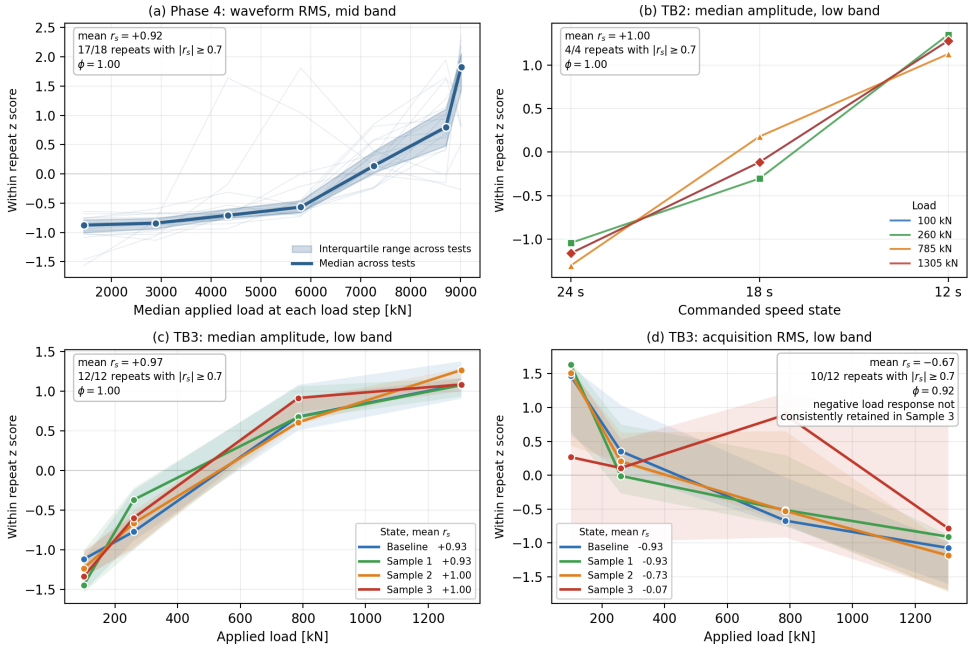


Fig. 3. Representative operational relationships, standardised within their own operating points. (a) Phase 4 mid-band waveform RMS versus load. (b) TB2 amplitude versus speed. (c) TB3 amplitude versus load across contamination stages. (d) TB3 acquisition RMS versus load, whose coupling collapses at the highest stage. Shaded regions show between-repeat spans and are not confidence intervals.

TABLE III
TB3 LOW-BAND LOAD CORRELATION PER CONTAMINATION STAGE
(MEAN OF THREE SPEED REPEATS)

Feature	Baseline	Sample 1	Sample 2	Sample 3
Amplitude	+0.93	+0.93	+1.00	+1.00
Duration / counts	−0.87	−0.87	−0.93	−0.87
Energy / duration	+0.87	+0.87	+0.93	+0.93
Energy	+0.93	+0.93	+1.00	+0.53
Counts	+0.93	+0.93	+1.00	+0.58
Duration	+0.84	+0.91	+0.93	+0.67
Acquisition RMS	−0.93	−0.93	−0.73	−0.07
Hit rate	−0.47	−0.53	−0.67	−0.87

their correction improves one channel while degrading another. Spectral descriptors repeat to within 3 % but are weakly associated. In TB2, rise time is perfectly rank ordered with speed yet differs by 46.5 % between repeated states.

Models fitted over 0–15 barg do not extrapolate to 20 and 25 barg, where flange leakage and manual pressure adjustment coincide with a change in the observed response. The large residuals at these states therefore mark the limit of the supported correction range and should not be interpreted as evidence of abnormal bearing condition.

D. Residualisation Shifts the Dominant Organisation

The intact reference provides the control result. In TB2, where no condition change occurs, three components explain 93.3 % of the variance. The retained scores are organised by speed-profile phase at $R^2_{\text{group}} = 0.360$ (Table V). Operating variation alone thus structures the joint feature space.

TABLE IV
CORRECTION ELIGIBILITY OUTCOMES. PREDICTIVE ADEQUACY IS THE CROSS-VALIDATED VARIANCE REDUCTION ON MID-BAND CHANNELS 2, 5 AND 8

Dataset	Feature (band)	Association	Stability	Repeat. (%)	Predictive adequacy (%)	Outcome
Swivel	Amplitude (mid)	Speed -0.82	✓	0.6	$+28$ / $+59$ / $+25$	**Accepted**
	Waveform RMS (mid)	Speed -0.95	✓	2.9	$+25$ / $+61$ / $+19$	**Accepted**
	Crest factor (mid)	Pressure $-0.77^{\ddagger}$	✓	3.3	$+22$ / $+45$ / -8	Fails predictive adequacy
	Kurtosis (mid)	Pressure $-0.81^{\ddagger}$	✓	9.0	-15 / $+83$ / -50	Fails predictive adequacy
	Hit rate per minute (mid)	Pressure $+0.88$	✓	80.7	not evaluated	Rejected at repeatability
	Hit rate per rotation (mid)	Pressure $+0.88$	✓	76.0	-3 / $+36$ / $+23$	Rejected at repeatability
	Spectral descriptors	< 0.7	–	≤ 2.9	not modelled	Repeatable, weakly associated
TB2	Amplitude (low)	Speed $+1.00$	✓	0.6	not testable	Supported
	Rise time (low)	Speed $+1.00$	✓	46.5	not testable	Rejected at repeatability
TB3	Amplitude (low)	Load $+0.97$	✓	not performed	not testable	Stable across stages
	Acquisition RMS (low)	Load -0.67	×	not performed	not testable	Rejected, coupling is stage dependent
Phase 4	Waveform RMS (mid)	Load $+0.92$	✓	not testable	not testable	Supported, robust to SNR filtering
	Hit rate (high)	Load $+0.89$	✓	not testable	not testable	Supported, population sensitive

Repeatability is the median d. Dash: fails the association screen. Swivel associations are 0–15 barg values, except ‡full sweep (-0.32 and -0.48 over 0–15 barg).

In raw TB3, four components explain $91.2\,\%$ of the variance, and the scores are organised more strongly by speed and phase than by stage. After baseline-fitted residualisation the ordering reverses (Fig. 4). Load, speed and phase association fall to 0.010, 0.060 and 0.073, and stage association rises from 0.269 to 0.468. The pattern holds for both sensitivity vectors and for retention thresholds of 80 and $95\,\%$ (0.493 and 0.459). Residual PC1 carries $56.8\,\%$ of the residual variance and contrasts low-band energy, counts and RMS against mid-band amplitude. Sample 3 is displaced along this component, while the baseline, Sample 1 and much of Sample 2 remain close together.

Several checks qualify the pooled values. First, the residual phase association is unevenly distributed over the components. It is only 0.021 on PC1 but 0.174 and 0.183 on PC3 and PC4, which carry $8.3\,\%$ and $5.4\,\%$ of the variance. $94\,\%$ of the stage association resides in PC1. Phase structure has thus been displaced into low-variance directions rather than eliminated.

Second, the stage response depends on speed. At matched load and phase combinations, Sample 3 is displaced in 17 of the 18 combinations with direct baseline support. Its median residual PC1 shift ranges, however, from -5.0 and -6.9 at the two lowest-speed phases to -22.1 at the highest-speed phase. Sample 2 is displaced most at the lower-speed phases and hardly at all at the highest, so the response is not monotonic in contamination level either.

TABLE V
SHARE OF RETAINED PCA SCORE VARIATION ASSOCIATED WITH EACH LABEL (R^2_{group})

Feature vector	Space	Load	Speed	Phase	Stage
TB2 core	Raw	$0.158^{\dagger}$	0.136	0.360	–
TB3 core	Raw	0.043	0.334	0.353	0.269
TB3 core	Residual	0.010	0.060	0.073	0.468
TB3 core, complete case	Residual	–	0.129	0.253	0.385
TB3 extended	Raw	–	0.386	0.404	0.206
TB3 extended	Residual	–	0.056	0.075	0.549
TB3 core + high band	Raw	–	0.278	0.294	0.352
TB3 core + high band	Residual	–	0.050	0.061	0.546

†Confounded with threshold retuning.

The reduction is also sensitive to sparse baseline support. Omitting the five imputed windows raises the residual phase association to 0.253, although stage remains the largest at 0.385.

Residualisation makes the highest contamination stage more distinct in the PCA score space, but does not improve the overall separation of all contamination stages. The stage silhouette decreases from 0.095 to 0.047, with the baseline and Sample 1 remaining unresolved.

V. DISCUSSION

A. Operational Sensitivity Is a Property of the Configuration

The results agree with earlier findings that speed is a first-order influence on AE and that the load response is less consistent [7], [8]. They extend them in two respects. First, the direction of the response is not a fixed property of a feature, as the opposite amplitude responses to speed in the two rigs show. The positive speed dependence of AE RMS reported under full-film conditions [7], [8] can therefore not be assumed in the friction-dominated regime at very low speed [2]. Second, a rank-order screen understates dependence when the response changes direction, as for the swivel bearing above $15\,\mathrm{barg}$. Several explanations for that change are plausible: a colder and more viscous grease in the high-pressure recordings, which were acquired first, a change in internal load distribution, or a shift of lubrication regime. They remain hypotheses, because pressure is confounded with test order and neither temperature nor friction torque was measured.

The contamination response differs from that reported by Miettinen and Andersson [9], since TB3 shows a higher event rate with little change in amplitude. The particles used here are coarser, up to about $500\,\mu\mathrm{m}$, and were introduced cumulatively, so the published relationship may not transfer, and the mechanism is not established.

B. From Association to Correction

Reproducibility, operational sensitivity and correctability proved to be distinct properties. Selecting correction targets by association strength alone would have admitted hit rate,

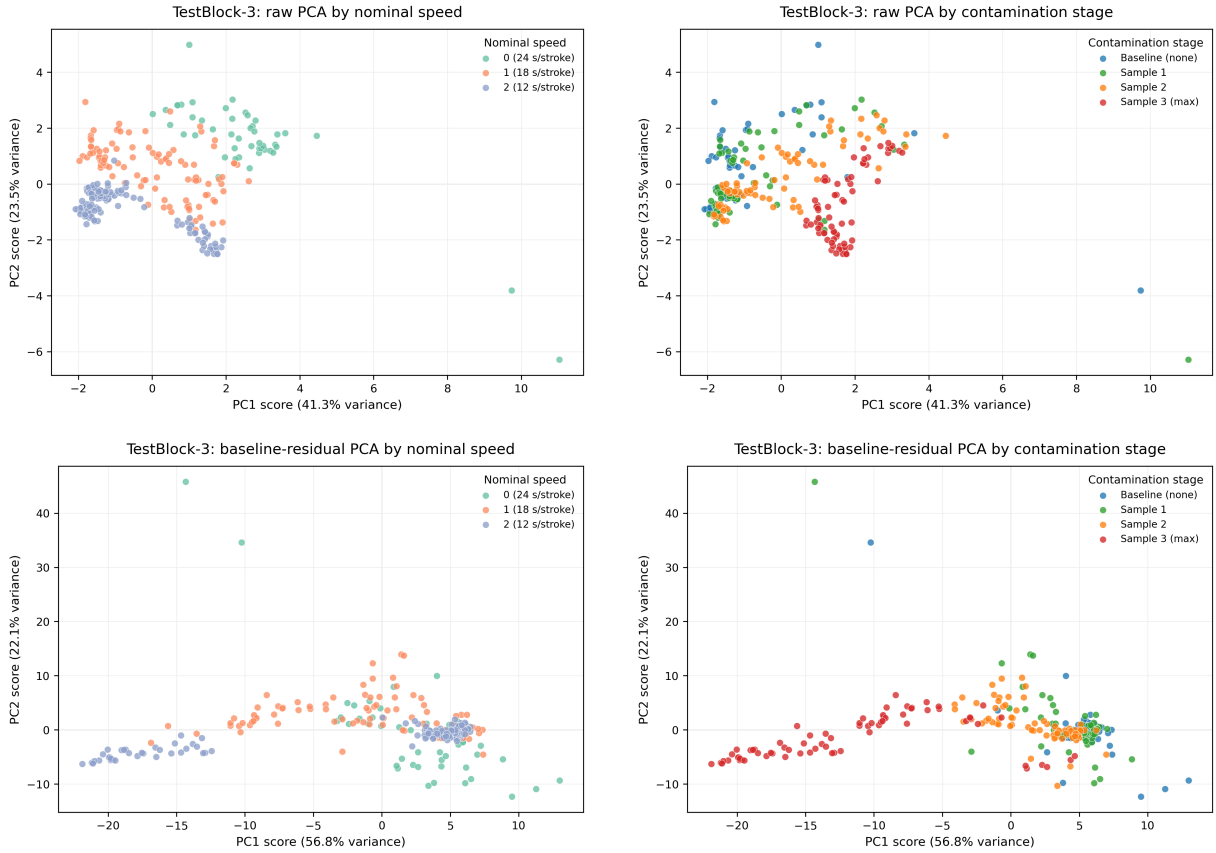


Fig. 4. TB3 PCA scores. From left to right and top to bottom: raw space coloured by nominal speed, raw space coloured by contamination stage, residual space coloured by nominal speed, and residual space coloured by contamination stage.

rise time, crest factor and kurtosis. The bounded correction domain is consistent with the observation of Worden *et al.* that a regression reference is valid only where the operating space has been sampled [14]. The present results add that the reference is also specific to the sensor, since a correction that improved one mid-band channel degraded nominally equivalent ones.

Of greater consequence, the operational relationship itself can depend on condition. When acquisition RMS loses its load coupling between the baseline and the final contamination stage, a residual contains both condition information and error from an outdated operational model. This bounds what a correction on load and speed can achieve for grease-lubricated bearings whose lubricant state is not measured.

C. Interpreting the Residual Feature Space

The TB2 control result shows that the central SHM caution also holds for AE. Multivariate structure in an AE feature space is not in itself evidence of a condition change [12], [13], [16]. Reducing the measured operational influence before the decomposition avoids having to decide which variance directions represent operating variability [15], and here it changed the dominant organisation from operating state to stage. The result is not circular, because the residual model was fitted on baseline windows only and PCA receives no stage labels.

The reduction is nevertheless partial, since residualisation removes the fitted conditional mean only. Operating-dependent variance and covariance [15], nonlinear dependence [17] and unmeasured influences remain. An additive model also cannot remove the observed stage-by-speed interaction by construction. A pooled association measure estimates a main effect and did not detect that interaction, which became visible only at matched operating states.

The threshold-like displacement of Sample 3 should not be read as a contamination threshold. A response that appears only above some severity, and a response that is present throughout but detectable only where the operating condition amplifies it, are equally compatible with the evidence. Residual PC1 is likewise a statistical feature contrast, and its attribution to a source mechanism would require waveform-level evidence and independent condition labels.

D. Implications for Offshore Bearing Maintenance

Offshore slewing and swivel bearings cannot readily be opened for inspection, so maintenance decisions rest on indirect indicators. A raw AE trend or a fixed alarm level is not a reliable one, because an intact bearing already produces structured AE variation under normal load and speed changes. Such an alarm would respond to the operating programme before it responds to the bearing.

The results instead favour a baseline-referenced procedure. A reference campaign is run once per bearing, while it is without known damage, and covers every load–speed combination to be monitored, after which the processing chain is frozen. Later campaigns repeat the programme and are compared with the baseline at matched load–speed combinations. Monitoring thus becomes a scheduled sweep instead of continuous acquisition, which suits slow-moving bearings. The result can be reported as one number per campaign, the displacement from the baseline in units of baseline spread.

In practice, operating state must be logged with the AE data, and threshold, gain and sensor coupling must stay fixed. Each bearing and sensor needs its own reference, and measurements outside the baseline range must be flagged as out of domain. A departure from the reference then triggers inspection or grease sampling but is not a diagnosis.

E. Limitations

The held-out correction and the condition-varying comparison come from different datasets, so neither half of the workflow validates the other. No dataset pairs a healthy and a mechanically damaged state on one rig, and only a late-stage contamination response is shown, so early warning is not demonstrated. The statistical evidence is descriptive, because correlations rest on four to seven operating points and observations are temporally dependent.

VI. Conclusion

Operating state is a first-order source of AE feature variability in low-speed rolling bearings. Its influence depends on feature definition, frequency band, sensor path, retained hit population and bearing condition. Association alone does not identify which features can be corrected, and stability, repeatability, range coverage and held-out predictive adequacy are also required. On a full-scale swivel bearing this sequence admitted only median amplitude and waveform RMS, over 0–$15\,\mathrm{barg}$.

The predictable operational component can then be reduced by low-order regression on a reference condition. In the contamination progression this shifted the dominant organisation of the PCA score space from operating state to contamination stage. The reduction was partial, as phase structure persisted in low-variance components, the stage response depended on speed, and the stage distributions continued to overlap.

For the maintenance of offshore bearings with limited inspection access, the quantity to monitor is the departure from a bearing's own operationally corrected reference at matched operating states rather than the raw feature level. This requires a baseline campaign per bearing, logged operating state and fixed acquisition settings. Such a departure indicates when inspection or grease sampling is warranted, but it does not identify the cause. The work does not provide a severity scale, a decision threshold or a validated classifier. These require healthy and naturally degrading states on one bearing with unchanged instrumentation, a measure of lubricant state, and correction models whose operational sensitivity may change with condition.

Acknowledgment

The linear bearing data were acquired within the HiTeAM joint industry programme at the Huisman Equipment R&D laboratory. The swivel bearing experiment was performed at the Bluewater Test Facility.

References

[1] B. Scheeren, M. L. Kaminski, and L. Pahlavan, "Acoustic emission monitoring of naturally developed damage in large-scale low-speed roller bearings," *Structural Health Monitoring*, vol. 23, no. 1, pp. 360–382, 2024.

[2] W. Caesarendra, B. Kosasih, A. K. Tieu, H. Zhu, C. A. S. Moodie, and Q. Zhu, "Acoustic emission-based condition monitoring methods: Review and application for low speed slew bearing," *Mechanical Systems and Signal Processing*, vol. 72–73, pp. 134–159, 2016.

[3] I. El-Thalji and E. Jantunen, "A descriptive model of wear evolution in rolling bearings," *Engineering Failure Analysis*, vol. 45, pp. 204–224, 2014.

[4] N. Jamaludin, D. Mba, and R. H. Bannister, "Condition monitoring of slow-speed rolling element bearings using stress waves," *Proceedings of the Institution of Mechanical Engineers, Part E: Journal of Process Mechanical Engineering*, vol. 215, no. 4, pp. 245–271, 2001.

[5] N. Tandon and B. C. Nakra, "Comparison of vibration and acoustic measurement techniques for the condition monitoring of rolling element bearings," *Tribology International*, vol. 25, no. 3, pp. 205–212, 1992.

[6] M. Elforjani and D. Mba, "Monitoring the onset and propagation of natural degradation process in a slow speed rolling element bearing with acoustic emission," *Journal of Vibration and Acoustics*, vol. 130, no. 4, p. 041013, 2008.

[7] J. Couturier and D. Mba, "Operational bearing parameters and acoustic emission generation," *Journal of Vibration and Acoustics*, vol. 130, no. 2, p. 024502, 2008.

[8] A. Cockerill, A. Clarke, R. Pullin, T. Bradshaw, P. Cole, and K. M. Holford, "Determination of rolling element bearing condition via acoustic emission," *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology*, vol. 230, no. 11, pp. 1377–1388, 2016.

[9] J. Miettinen and P. Andersson, "Acoustic emission of rolling bearings lubricated with contaminated grease," *Tribology International*, vol. 33, no. 11, pp. 777–787, 2000.

[10] B. Scheeren, N. P. Thakoerdajal, and L. Pahlavan, "Influence of grease contamination on acoustic emission monitoring of low-speed roller bearings," *e-Journal of Nondestructive Testing*, pp. 1–8, 2024.

[11] Y. Guo, A. Thomsen, R. Bergua, O. Bankestrom, J. Erskine, and J. Keller, "Acoustic emission measurement of a wind turbine main bearing," National Renewable Energy Laboratory, Golden, CO, Tech. Rep. NREL/TP-5000-83370, Oct. 2022.

[12] H. Sohn, "Effects of environmental and operational variability on structural health monitoring," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 365, no. 1851, pp. 539–560, 2007.

[13] E. Figueiredo and J. Brownjohn, "Three decades of statistical pattern recognition paradigm for SHM of bridges," *Structural Health Monitoring*, vol. 21, no. 6, pp. 3018–3054, 2022.

[14] K. Worden, H. Sohn, and C. R. Farrar, "Novelty detection in a changing environment: regression and interpolation approaches," *Journal of Sound and Vibration*, vol. 258, no. 4, pp. 741–761, 2002.

[15] L. Neumann, P. Wittenberg, and J. Gertheiss, "Removal of multivariate environmental influences in structural health monitoring through conditional covariances and supervised learning," arXiv:2606.05026 [stat.ME], 2026.

[16] A. Widodo, J.-D. Son, B.-S. Yang, Y.-H. Kim, A. C. C. Tan, J. Mathew, D.-S. Gu, and B.-K. Choi, "Fault diagnosis of low speed bearing based on acoustic emission signal and multi-class relevance vector machine," in *Proceedings of the 15th International Congress on Sound and Vibration (ICSV15)*, Daejeon, Korea, 2008, pp. 1468–1475.

[17] A.-M. Yan, G. Kerschen, P. De Boe, and J.-C. Golinval, "Structural damage diagnosis under varying environmental conditions—part II: local PCA for non-linear cases," *Mechanical Systems and Signal Processing*, vol. 19, pp. 865–880, 2005.

[18] B. Scheeren and L. Pahlavan, "Report phase 4: HiTeAM – high-tech acoustic monitoring," HiTeAM Joint Industry Programme, Tech. Rep., 2025, version 1.0.

[19] B. Scheeren, A. Huijer, F. Riccioli, N. Thakoerdajal, and P. Pahlavan, "Report phase 2: HiTeAM – high-tech acoustic monitoring," HiTeAM Joint Industry Programme, Tech. Rep., Dec. 2021, version 2.0.
[20] Vallen Systeme GmbH, *AMSY-6 System Description*, Vallen Systeme GmbH, Wolfratshausen, Germany, Nov. 2021.
[21] P. J. Rousseeuw and M. Hubert, "Anomaly detection by robust statistics," *WIREs Data Mining and Knowledge Discovery*, vol. 8, p. e1236, 2018.
[22] I. T. Jolliffe, *Principal Component Analysis*, 2nd ed. New York: Springer, 2002.
[23] M. J. Anderson, "A new method for non-parametric multivariate analysis of variance," *Austral Ecology*, vol. 26, no. 1, pp. 32–46, 2001.

B

Linear Bearing: Supporting Data

This appendix contains the supporting data behind the cross-test summaries of chapter 4: the sensor configurations, the recorded hit populations, the per-test load-sensitivity matrices, the evidence underlying two data-quality findings, the speed-sensitivity summary and the lubrication results.

Spearman correlations near $+1$ indicate a monotonic increase with load or speed, near -1 a monotonic decrease, and near 0 weak or inconsistent monotonic dependence. Cells with $|r_s| \geq 0.7$ are set in bold. Because each coefficient rests on a small number of load or speed steps, the values are interpreted as trend indicators rather than as precise effect sizes. A bold cell therefore marks only an association that clears the screening criterion of section 3.5 and is not by itself evidence of correction eligibility. Eligibility follows the criteria of section 3.6 and is reported in sections 4.2.7, 4.3.6 and 4.4.7.

The appendix provides supporting evidence rather than a complete record of every intermediate product. Two products are summarised rather than reproduced. The per-test waveform-level matrices are represented by the single triplet read individually in the chapter (section B.5), and the per-triplet SNR-filtered matrices by their cross-test summaries (section B.6). In both cases the conclusions of chapter 4 rest on the cross-test behaviour rather than on individual matrices.

B.1. Sensor and Channel Configuration

In both linear bearing rigs each sensor node carries three transducers, and the corresponding acquisition channels are band-pass filtered to the low, mid and high bands defined in section 4.2.3. The band limits are acquisition settings rather than transducer operating ranges.

Table B.1 gives the Phase 4 mapping over six sensor nodes and eighteen channels. Node positions are shown in figure 4.1: A–B are drive-end raceway locations, C–D bearing-body locations and E–F far-end raceway locations.

Location	Low ($40\,\mathrm{kHz}$ to $100\,\mathrm{kHz}$)	Mid ($95\,\mathrm{kHz}$ to $180\,\mathrm{kHz}$)	High ($180\,\mathrm{kHz}$ to $580\,\mathrm{kHz}$)
A	1	3	5
B	2	4	6
C	7	9	11
D	8	10	12
E	13	15	17
F	14	16	18

Table B.1: Phase 4 channel matrix.

Table B.2 gives the Phase 2 mapping, shared by TestBlock 2 and TestBlock 3, over eight locations and 24 channels. Channels 1–12 lie in the lower chamber and 13–24 in the upper. Within each chamber, odd-numbered channels are on the support structure and even-numbered channels on the nose ring.

Location	Low ($40\,\mathrm{kHz}$ to $100\,\mathrm{kHz}$)	Mid ($95\,\mathrm{kHz}$ to $180\,\mathrm{kHz}$)	High ($180\,\mathrm{kHz}$ to $580\,\mathrm{kHz}$)
L-Raceway-Support	1	3	5
L-Raceway-Nose	2	4	6
L-Substructure-Support	7	9	11
L-Substructure-Nose	8	10	12
U-Substructure-Support	13	15	17
U-Substructure-Nose	14	16	18
U-Raceway-Support	19	21	23
U-Raceway-Nose	20	22	24

Table B.2: Phase 2 channel matrix.

Channels 1–6 and 19–24 are raceway locations and 7–18 substructure locations [33].

B.2. Operational Characterisation Sequence

Figure B.1 shows the sequence applied to each dataset in chapter 4. Hits are assigned an operating state and aggregated within the relevant operating step. Association is screened first, followed by stability, repeatability where a repeated state exists, and range coverage. Criteria that a design cannot support are reported as not testable rather than assumed to be met.

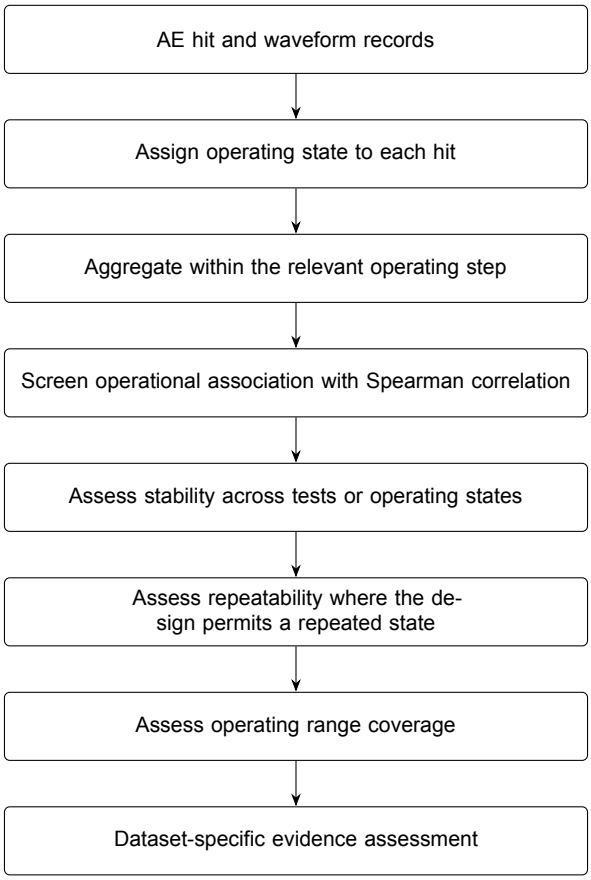


Figure B.1: Operational characterisation sequence.

B.3. Recorded Data Overview

This section describes what was recorded, before any screening or operational interpretation. The figures show the hit population as stored in the acquisition databases, so that the analytical populations

of chapter 4 can be read against the material they were drawn from.

The three datasets differ most in how their activity is distributed, and that difference governs what each can support. Table B.3 gives the distribution across frequency bands together with the share carried by the two busiest channels, which measures how far a dataset behaves as a single-location measurement. The TestBlock 3 figures exclude a separate pulsing calibration recording.

Dataset	Recorded hits	Low	Mid	High	Two busiest channels
Phase 4	7 899 057	63.41 %	22.45 %	14.14 %	36.23 %
TestBlock 2	1 909 077	99.28 %	0.66 %	0.06 %	99.16 %
TestBlock 3	5 383 227	96.16 %	2.70 %	1.14 %	90.56 %

Table B.3: Recorded hit population by dataset.

Two consequences follow. Phase 4 populates all three bands and spreads its activity across the sensor array, so its results can be reported band specifically and are not a property of one sensor position. Both Phase 2 blocks are dominated by the low band and by the same two support-mounted channels, so neither supports an assessment of sensor-position dependence. The two Phase 2 blocks also differ in degree. TestBlock 2 retains only tens to hundreds of mid- and high-band hits per load–speed combination, which is too few for a stable median, whereas TestBlock 3 retains approximately 145 000 mid-band and 61 000 high-band hits in total. TestBlock 2 is therefore reported as a low-band dataset and TestBlock 3 across all three bands.

B.3.1. Phase 4

Figure B.2 shows activity distributed across all six sensor nodes and all three bands, with the busiest channel carrying about one fifth of the total. This distribution permits the band-specific reporting used throughout section 4.2. In figure B.3 the population rises from the acquisition threshold without a sharp cliff, and the three bands occupy overlapping amplitude ranges. Figure B.4 shows that recording sizes vary by more than an order of magnitude across the programme, which is why cross-test comparisons are made on correlations rather than on absolute activity. Figure B.5 shows event durations spanning several decades, which is why per-hit features are aggregated by median rather than by mean.

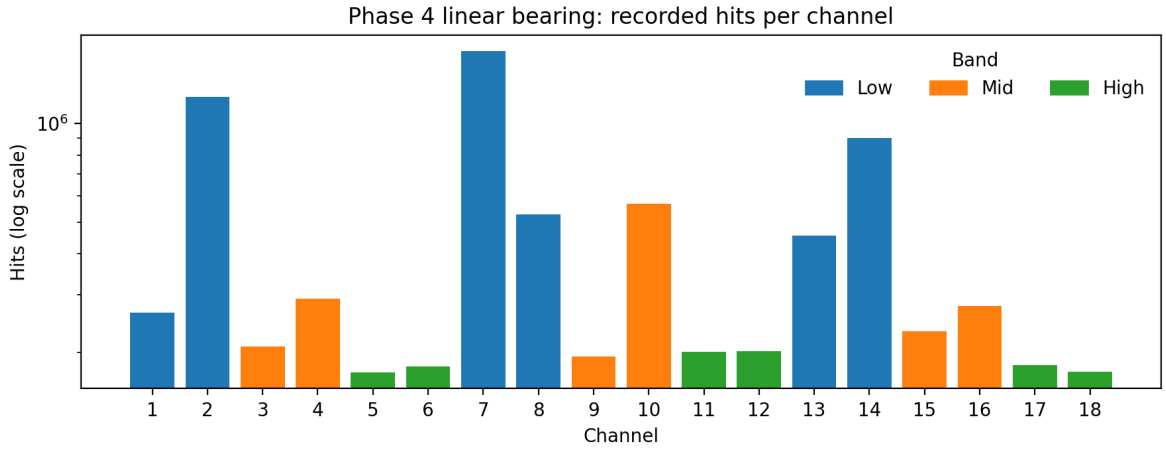


Figure B.2: Phase 4 hits per channel by band.

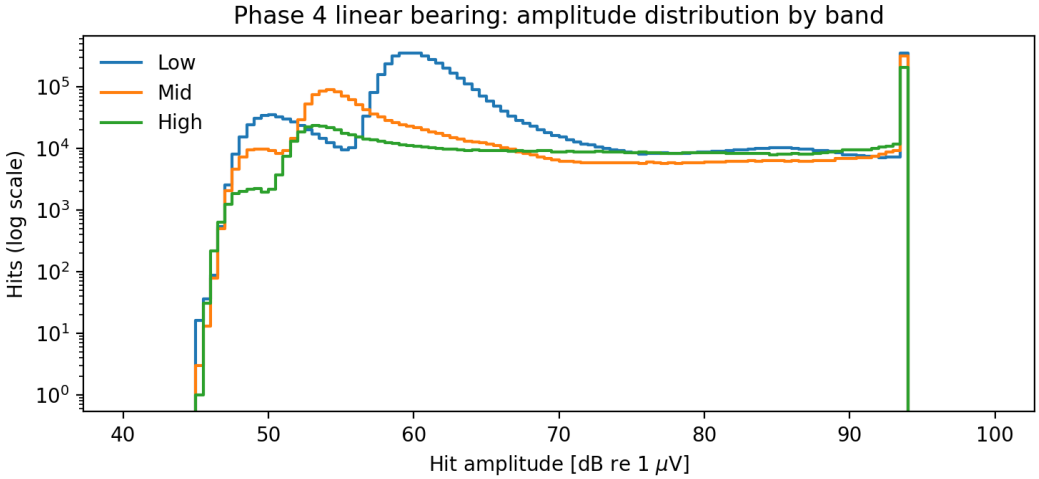


Figure B.3: Phase 4 amplitude distribution by band.

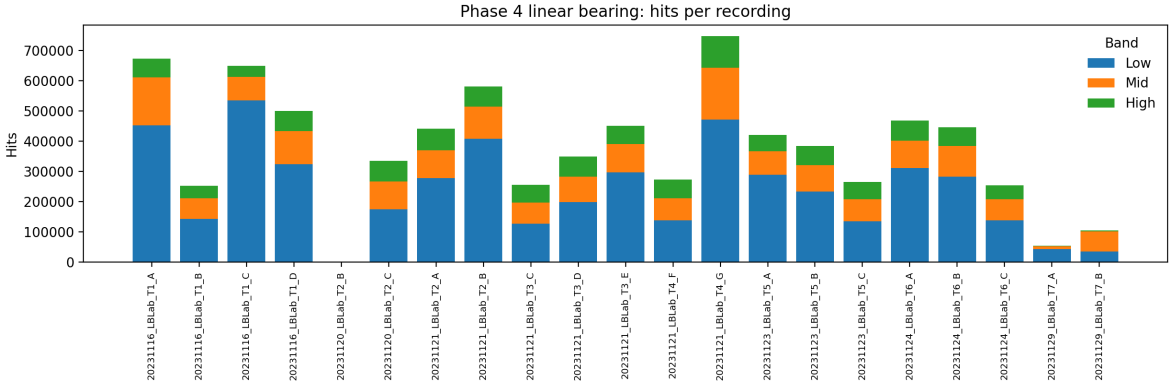


Figure B.4: Phase 4 hits per acquisition file.

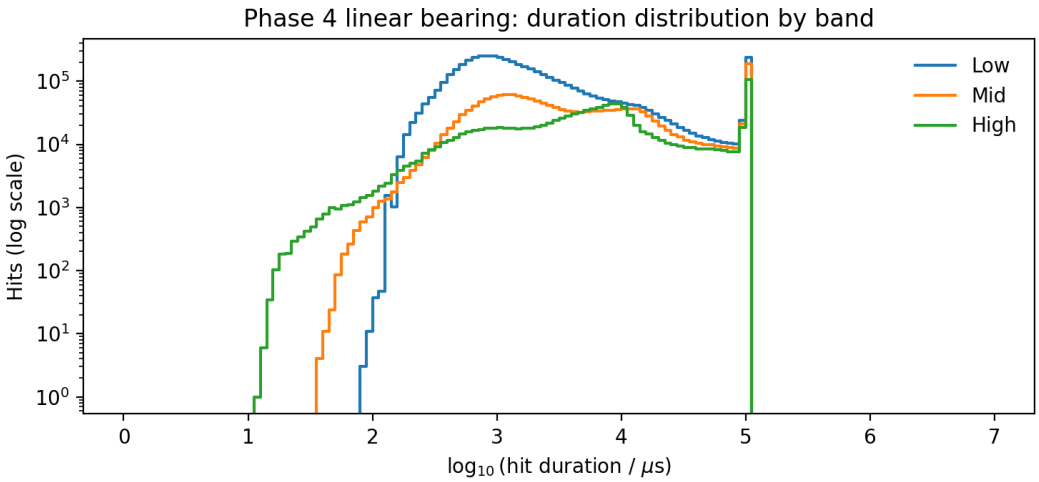


Figure B.5: Phase 4 duration distribution by band.

B.3.2. TestBlock 2

Figure B.6 uses a logarithmic vertical scale. Channels 1 and 7, both low-frequency support-mounted sensors in the lower chamber, carry $99.16\,\%$ of the block, and every other channel is two to four orders of magnitude quieter. This concentration makes TestBlock 2 effectively a single-band, single-location measurement.

In figure B.7 the low-band population begins at the acquisition floor and rises without a cliff, which is why that floor rather than the Phase 4 level is used as the screening reference in section 4.3.4. Because the threshold on channels 1 and 7 was retuned between recordings (table 4.6), the aggregated distribution superimposes populations recorded at different detection thresholds.

Each file in figure B.8 is one load step. Recorded activity falls across the sequence, which section 4.3.3 attributes to the concurrent threshold increase rather than to load. In figure B.9 the mid- and high-band populations are too sparse to define a distribution, the same limitation that prevents a stable median per load–speed combination in those bands.

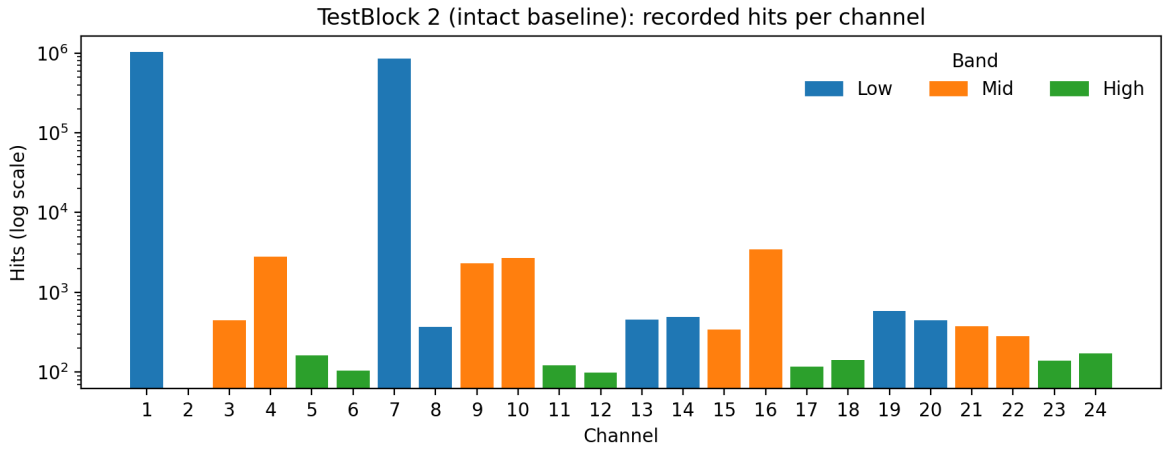


Figure B.6: TestBlock 2 hits per channel by band (logarithmic scale).

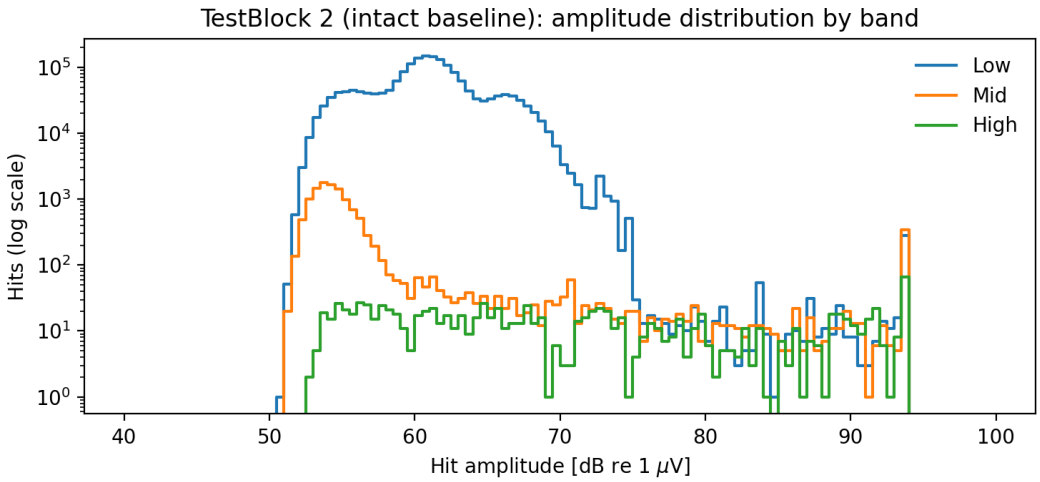


Figure B.7: TestBlock 2 amplitude distribution by band.

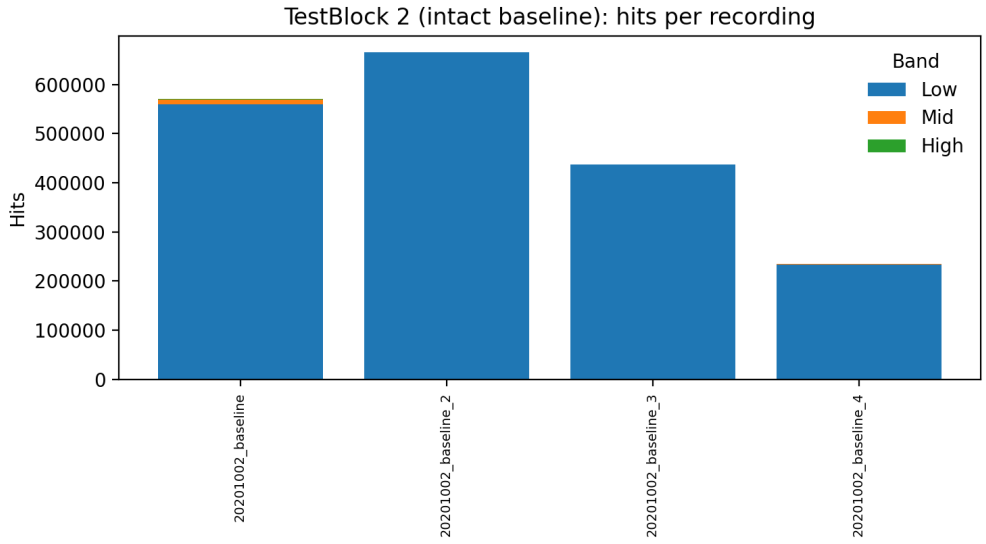


Figure B.8: TestBlock 2 hits per acquisition file.

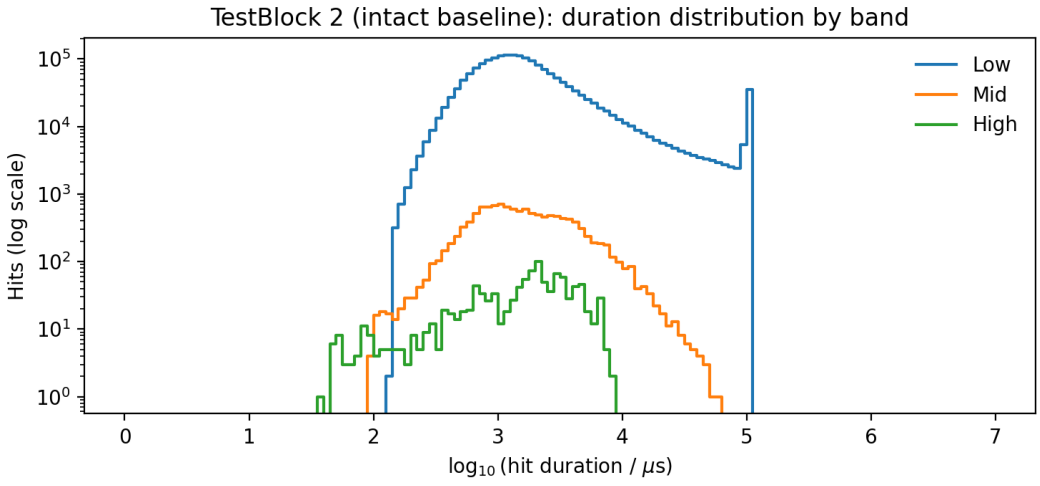


Figure B.9: TestBlock 2 duration distribution by band.

B.3.3. TestBlock 3

Figure B.10 shows the same concentration on channels 1 and 7 as TestBlock 2, at $90.56\,\%$ rather than $99.16\,\%$, so TestBlock 3 is likewise close to a single-location measurement. The mid- and high-band channels are, however, populated well enough to support medians per load–speed combination.

Unlike figure B.7, the low-band population in figure B.11 has a single well-defined lower edge, consistent with a detection threshold that is stable across all four contamination runs. The fourteen contamination files of figure B.12 do not map one to one onto operating states, since several combine two or three load steps; the mapping is given in section B.8. In figure B.13 the low-band distribution terminates at a single repeated upper value rather than tapering. This is the acquisition censoring identified in section 4.4.2 and the reason low-band duration is excluded from the primary feature vector of chapter 6.

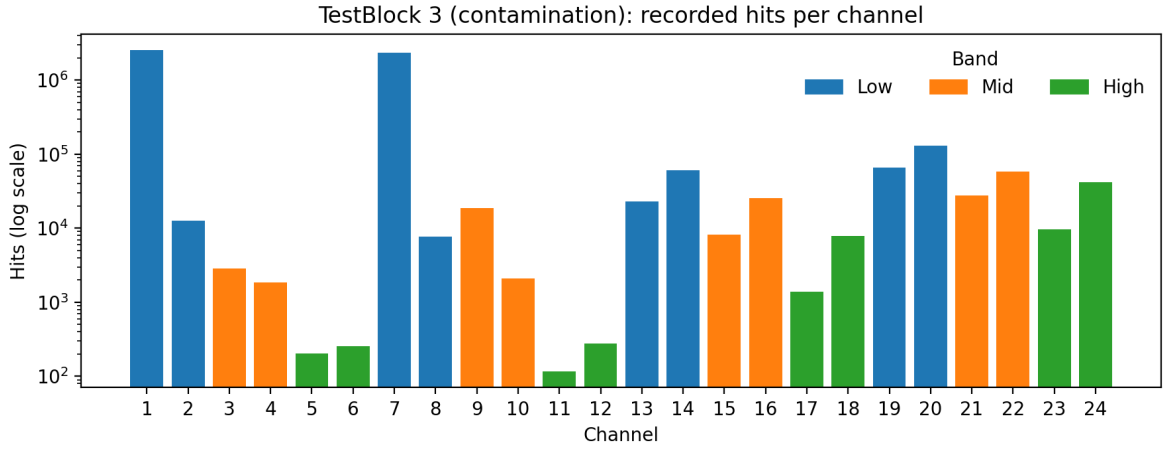


Figure B.10: TestBlock 3 hits per channel by band.

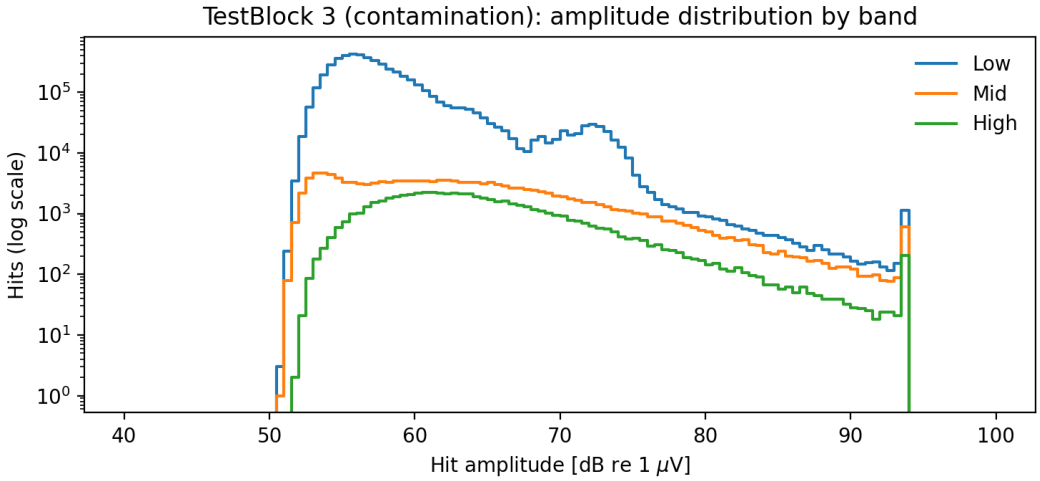


Figure B.11: TestBlock 3 amplitude distribution by band.

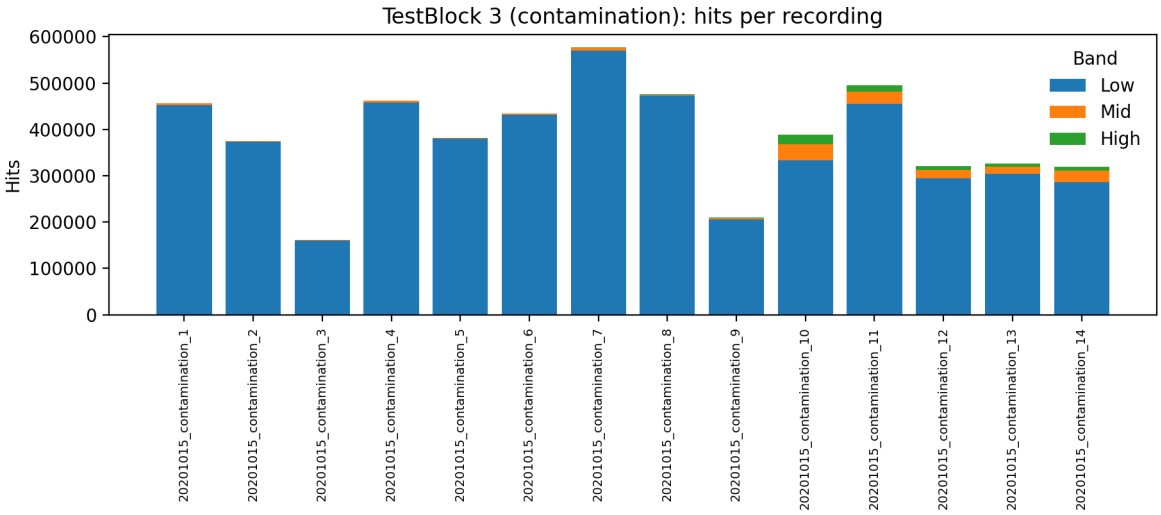


Figure B.12: TestBlock 3 hits per acquisition file.

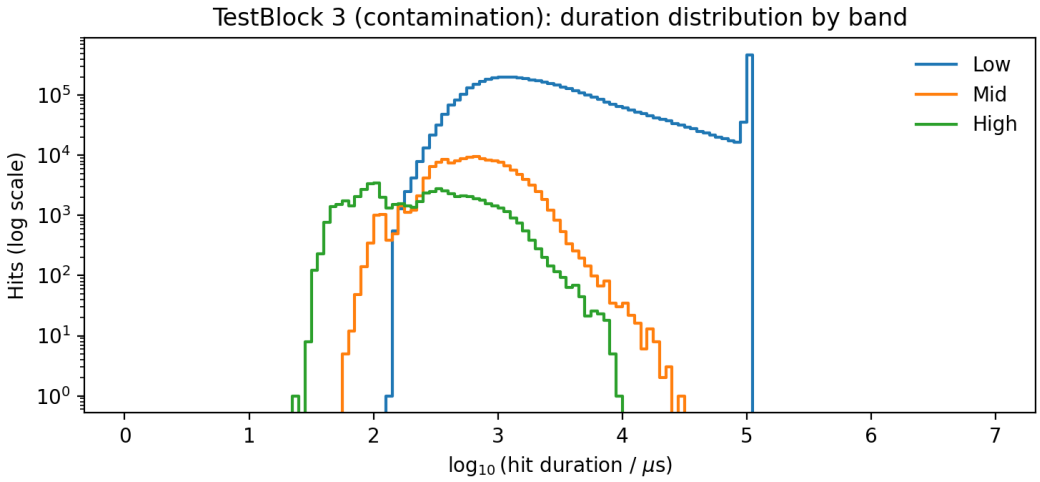


Figure B.13: TestBlock 3 duration distribution by band.

B.4. Scalar Load-Sensitivity Matrices

Tables B.4 to B.6 contain the per-test matrices behind the cross-test means of table 4.3. Each cell is the Spearman correlation between one aggregated acquisition quantity or derived ratio (equation (3.6)) and the measured force F_z, computed across the load steps of one test, and bold marks $|r_s| \geq 0.7$.

The eighteen tests are grouped by lubrication condition, three tests per sensor node position, with stroke speed increasing from left to right within each node. Reading a row across its six columns gives the consistency of one feature–band relationship across the repeats that table 4.3 averages over. The bracketed counts in that table are the number of bold cells in the corresponding row, summed across the three groups. The 3-2 traveller-node triplet in table B.6 is the triplet discussed in section 4.2.4.

The quantity labelled *RMS (acq.)* is $\mathrm{RMS}_{\mathrm{acq}}$, the acquisition-supplied RMS of the background noise preceding each threshold crossing, which excludes the hit signal itself (section 4.2.4). It therefore describes the measurement background at the moment of detection rather than the detected transient. It is a different measurement from the waveform RMS of section B.5, which is recomputed over the whole stored transient, and the two are not expected to agree.

Quantity	Band	Raceway node (1-1)			Traveller node (1-2)		
		1-1-1	1-1-2	1-1-3	1-2-1	1-2-2	1-2-3
		0.095	0.140	0.165	0.095	0.140	0.165
		stroke speed [$\mathrm{m\,s^{-1}}$]					
Hit-rate	Low	-0.14	$+0.61$	$\mathbf{+0.96}$	-0.25	-0.07	$+0.61$
	Mid	-0.21	$\mathbf{+0.79}$	$\mathbf{+0.93}$	$+0.11$	$+0.46$	$\mathbf{+0.89}$
	High	$+0.54$	$\mathbf{+0.96}$	$\mathbf{+0.96}$	$\mathbf{+0.96}$	$\mathbf{+0.89}$	$\mathbf{+0.96}$
Duration	Low	$+0.11$	$\mathbf{+0.89}$	$\mathbf{+0.89}$	$\mathbf{+0.86}$	$+0.68$	$\mathbf{+1.00}$
	Mid	$+0.43$	$\mathbf{+0.93}$	$+0.39$	$\mathbf{+0.96}$	$\mathbf{+1.00}$	$\mathbf{+0.96}$
	High	$+0.21$	$+0.61$	$+0.43$	$\mathbf{+0.89}$	-0.07	-0.11
RMS (acq.)	Low	-0.64	$\mathbf{-0.82}$	-0.54	$\mathbf{-0.86}$	$\mathbf{-0.75}$	$\mathbf{+0.71}$
	Mid	$\mathbf{-0.75}$	-0.32	$\mathbf{+0.93}$	-0.43	-0.32	$\mathbf{+0.93}$
	High	$+0.43$	$\mathbf{+0.93}$	$\mathbf{+0.89}$	$\mathbf{+1.00}$	$\mathbf{+0.96}$	$\mathbf{+0.96}$
Energy	Low	-0.07	$\mathbf{+0.86}$	$+0.50$	$+0.61$	$+0.54$	$\mathbf{+1.00}$
	Mid	$+0.64$	$\mathbf{+0.96}$	$+0.39$	$\mathbf{+0.96}$	$\mathbf{+1.00}$	$\mathbf{+0.96}$
	High	$+0.21$	$+0.57$	$+0.39$	$\mathbf{+0.96}$	$+0.43$	$+0.57$
Amplitude [dB]	Low	-0.07	$\mathbf{+0.86}$	$+0.61$	$+0.43$	$+0.43$	$\mathbf{+0.96}$
	Mid	$\mathbf{+0.75}$	$\mathbf{+0.96}$	$+0.25$	$\mathbf{+0.96}$	$\mathbf{+1.00}$	$\mathbf{+0.96}$
	High	$+0.36$	$+0.61$	$+0.14$	$\mathbf{+0.96}$	$\mathbf{+0.79}$	$\mathbf{+0.71}$
Rise time	Low	$+0.07$	$\mathbf{+0.96}$	$\mathbf{+0.71}$	$\mathbf{+0.89}$	$+0.54$	$\mathbf{+0.96}$
	Mid	-0.61	$\mathbf{-0.96}$	$+0.43$	-0.61	-0.61	$+0.46$
	High	$+0.18$	$+0.25$	$+0.43$	$+0.43$	-0.61	$\mathbf{-0.71}$
Counts	Low	-0.02	$\mathbf{+0.88}$	$+0.67$	$+0.66$	$+0.54$	$\mathbf{+1.00}$
	Mid	$\mathbf{+0.70}$	$\mathbf{+0.96}$	$+0.39$	$\mathbf{+0.96}$	$\mathbf{+1.00}$	$\mathbf{+0.96}$
	High	$+0.21$	$+0.57$	$+0.43$	$\mathbf{+0.96}$	-0.07	-0.39
Duration / counts	Low	$+0.64$	$\mathbf{-0.75}$	$\mathbf{+0.75}$	$\mathbf{-0.82}$	-0.11	-0.57
	Mid	$\mathbf{-0.75}$	$\mathbf{-1.00}$	-0.39	$\mathbf{-0.96}$	$\mathbf{-1.00}$	$\mathbf{-0.89}$
	High	-0.18	-0.46	-0.39	$\mathbf{-0.96}$	$+0.07$	$+0.64$
Energy / duration	Low	$\mathbf{-0.71}$	$\mathbf{+0.89}$	-0.43	$+0.54$	$+0.29$	$+0.61$
	Mid	$\mathbf{+0.75}$	$\mathbf{+0.96}$	$+0.25$	$\mathbf{+0.96}$	$\mathbf{+1.00}$	$\mathbf{+0.89}$
	High	$+0.21$	$+0.61$	$+0.14$	$\mathbf{+0.96}$	$\mathbf{+0.79}$	$\mathbf{+0.71}$

Table B.4: Phase 4 scalar load sensitivity, starvation (groups 1-1 and 1-2).

Quantity	Band	Raceway node (2-1)			Traveller node (2-2)		
		2-1-1	2-1-2	2-1-3	2-2-1	2-2-2	2-2-3
		0.095	0.140	0.165	0.095	0.140	0.165
		stroke speed [$\mathrm{m\,s^{-1}}$]					
Hit-rate	Low	-0.46	$+0.46$	$\mathbf{+0.96}$	-0.43	$\mathbf{+0.96}$	$\mathbf{+0.96}$
	Mid	-0.43	$+0.54$	$\mathbf{+0.96}$	$+0.11$	$\mathbf{+1.00}$	$\mathbf{+0.96}$
	High	$\mathbf{+0.75}$	$+0.36$	$\mathbf{+0.93}$	$\mathbf{+0.96}$	$\mathbf{+1.00}$	$\mathbf{+1.00}$
Duration	Low	$+0.50$	$\mathbf{+0.89}$	$\mathbf{+0.93}$	$+0.39$	$+0.36$	$\mathbf{+0.93}$
	Mid	$\mathbf{+0.93}$	$+0.14$	$\mathbf{+0.71}$	$\mathbf{+0.96}$	$\mathbf{+0.93}$	$+0.00$
	High	$+0.04$	-0.64	-0.43	-0.04	-0.54	$\mathbf{-0.75}$
RMS (acq.)	Low	$\mathbf{-0.82}$	$\mathbf{-0.86}$	$+0.29$	$\mathbf{-0.75}$	-0.36	-0.29
	Mid	-0.32	$+0.25$	$+0.46$	-0.32	-0.50	$+0.54$
	High	$\mathbf{+0.96}$	$+0.32$	$+0.64$	$\mathbf{+0.93}$	$\mathbf{+0.89}$	$+0.61$
Energy	Low	$+0.54$	$\mathbf{+0.89}$	$\mathbf{+1.00}$	$+0.39$	$+0.18$	$\mathbf{+0.96}$
	Mid	$\mathbf{+1.00}$	$+0.50$	$\mathbf{+0.82}$	$\mathbf{+0.96}$	$\mathbf{+0.93}$	$+0.54$
	High	$+0.04$	$+0.43$	$+0.54$	$\mathbf{+0.86}$	$+0.43$	-0.14
Amplitude [dB]	Low	$+0.68$	$\mathbf{+0.89}$	$\mathbf{+0.89}$	$+0.39$	$+0.45$	$\mathbf{+1.00}$
	Mid	$\mathbf{+0.96}$	$\mathbf{+0.75}$	$\mathbf{+0.89}$	$\mathbf{+0.96}$	$\mathbf{+0.96}$	$\mathbf{+0.75}$
	High	$+0.61$	$\mathbf{+0.96}$	$\mathbf{+0.86}$	$\mathbf{+0.96}$	$\mathbf{+0.89}$	$+0.68$
Rise time	Low	$+0.39$	$+0.11$	$+0.43$	$+0.39$	-0.18	-0.25
	Mid	-0.64	-0.68	$\mathbf{-0.71}$	$\mathbf{-0.71}$	$\mathbf{-0.86}$	$\mathbf{-0.79}$
	High	-0.50	-0.64	-0.61	$\mathbf{-0.86}$	$\mathbf{-0.82}$	$\mathbf{-0.89}$
Counts	Low	$+0.50$	$\mathbf{+0.85}$	$\mathbf{+0.93}$	$+0.39$	$+0.53$	$\mathbf{+0.95}$
	Mid	$\mathbf{+0.96}$	$+0.54$	$\mathbf{+0.82}$	$\mathbf{+0.96}$	$\mathbf{+0.93}$	$+0.11$
	High	$+0.04$	-0.64	-0.21	-0.04	-0.54	$\mathbf{-0.79}$
Duration / counts	Low	$\mathbf{-0.89}$	-0.36	$\mathbf{-0.96}$	-0.68	-0.29	$\mathbf{-0.93}$
	Mid	$\mathbf{-1.00}$	-0.14	$\mathbf{-0.71}$	$\mathbf{-0.96}$	$\mathbf{-0.93}$	$+0.00$
	High	-0.04	$+0.57$	$+0.21$	-0.36	$+0.54$	$\mathbf{+0.79}$
Energy / duration	Low	$\mathbf{+0.89}$	$\mathbf{+0.79}$	$\mathbf{+0.96}$	$+0.68$	$+0.32$	$\mathbf{+0.96}$
	Mid	$\mathbf{+0.96}$	$+0.54$	$\mathbf{+0.82}$	$\mathbf{+0.96}$	$\mathbf{+0.96}$	$\mathbf{+0.75}$
	High	$+0.61$	$+0.64$	$\mathbf{+0.86}$	$\mathbf{+0.96}$	$\mathbf{+0.89}$	$+0.54$

Table B.5: Phase 4 scalar load sensitivity, full-mixed grease (groups 2-1 and 2-2).

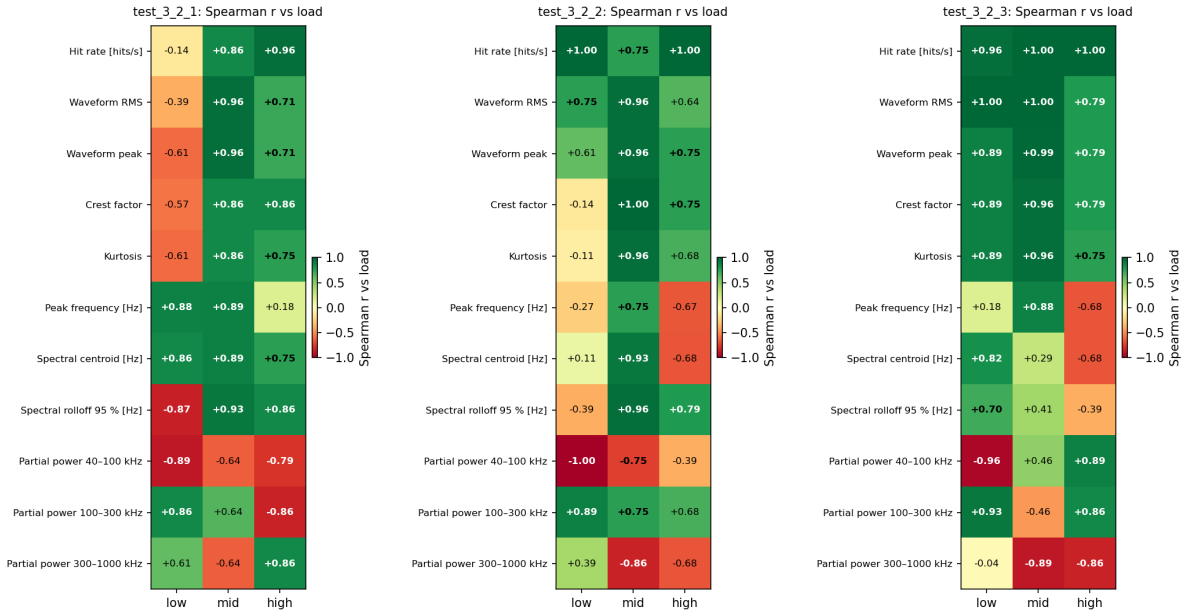
Quantity	Band	Raceway node (3-1)			Traveller node (3-2)		
		3-1-1	3-1-2	3-1-3	3-2-1	3-2-2	3-2-3
		0.095	0.140	0.165	0.095	0.140	0.165
		stroke speed [$\mathrm{m\,s^{-1}}$]					
Hit-rate	Low	$+0.32$	$\mathbf{+0.89}$	$\mathbf{+0.89}$	-0.14	$\mathbf{+1.00}$	$\mathbf{+0.96}$
	Mid	$+0.68$	$+0.57$	$\mathbf{+0.96}$	$\mathbf{+0.86}$	$\mathbf{+0.75}$	$\mathbf{+1.00}$
	High	$\mathbf{+0.89}$	$\mathbf{+1.00}$	$\mathbf{+0.96}$	$\mathbf{+0.96}$	$\mathbf{+1.00}$	$\mathbf{+1.00}$
Duration	Low	$+0.43$	$+0.64$	$\mathbf{+0.96}$	-0.11	$\mathbf{+0.75}$	$\mathbf{+0.96}$
	Mid	$\mathbf{+0.79}$	$\mathbf{+0.96}$	$\mathbf{+0.89}$	$+0.68$	$\mathbf{+0.93}$	$\mathbf{+0.86}$
	High	-0.07	$\mathbf{-0.75}$	-0.32	-0.68	-0.04	-0.29
RMS (acq.)	Low	-0.61	$\mathbf{-0.82}$	$\mathbf{-1.00}$	-0.61	$\mathbf{-1.00}$	$\mathbf{-0.93}$
	Mid	-0.43	-0.57	-0.14	-0.32	-0.64	-0.64
	High	$\mathbf{+0.96}$	$\mathbf{+0.96}$	$\mathbf{+0.96}$	$\mathbf{+0.96}$	$\mathbf{+0.96}$	$\mathbf{+0.71}$
Energy	Low	$+0.50$	$\mathbf{+0.89}$	$\mathbf{+1.00}$	-0.11	$\mathbf{+0.75}$	$\mathbf{+0.96}$
	Mid	$\mathbf{+0.96}$	$\mathbf{+1.00}$	$\mathbf{+0.96}$	$\mathbf{+0.96}$	$\mathbf{+0.96}$	$\mathbf{+0.93}$
	High	$+0.36$	$+0.29$	$+0.61$	$+0.11$	$\mathbf{+0.79}$	$+0.04$
Amplitude [dB]	Low	$+0.04$	$\mathbf{+0.75}$	$\mathbf{+1.00}$	-0.04	$+0.43$	$\mathbf{+0.96}$
	Mid	$\mathbf{+1.00}$	$\mathbf{+0.96}$	$\mathbf{+1.00}$	$\mathbf{+0.96}$	$\mathbf{+0.96}$	$\mathbf{+0.96}$
	High	$\mathbf{+0.79}$	$\mathbf{+1.00}$	$\mathbf{+0.96}$	$\mathbf{+0.96}$	$\mathbf{+0.96}$	$+0.57$
Rise time	Low	$+0.29$	$+0.14$	$\mathbf{+0.79}$	-0.07	$+0.57$	$\mathbf{+0.71}$
	Mid	$\mathbf{-0.93}$	$\mathbf{-0.79}$	-0.46	$\mathbf{-0.82}$	-0.68	$\mathbf{-0.75}$
	High	$\mathbf{-0.86}$	$\mathbf{-0.93}$	$\mathbf{-0.79}$	$\mathbf{-0.86}$	-0.64	-0.68
Counts	Low	$+0.58$	$\mathbf{+0.96}$	$\mathbf{+0.99}$	-0.09	$\mathbf{+0.78}$	$\mathbf{+1.00}$
	Mid	$\mathbf{+0.86}$	$\mathbf{+0.99}$	$\mathbf{+0.96}$	$\mathbf{+0.86}$	$\mathbf{+0.96}$	$\mathbf{+0.93}$
	High	$+0.36$	$\mathbf{-0.82}$	-0.57	-0.57	-0.29	-0.64
Duration / counts	Low	$\mathbf{-1.00}$	$\mathbf{-0.96}$	$\mathbf{-1.00}$	-0.18	-0.68	$\mathbf{-0.86}$
	Mid	$\mathbf{-1.00}$	$\mathbf{-0.89}$	$\mathbf{-0.96}$	$\mathbf{-1.00}$	$\mathbf{-0.96}$	$\mathbf{-0.96}$
	High	-0.36	$\mathbf{+0.82}$	$\mathbf{+0.75}$	$+0.54$	$+0.61$	$+0.64$
Energy / duration	Low	$\mathbf{+0.89}$	$\mathbf{+1.00}$	$\mathbf{+1.00}$	-0.11	$+0.54$	$\mathbf{+0.93}$
	Mid	$\mathbf{+1.00}$	$\mathbf{+0.89}$	$\mathbf{+0.96}$	$\mathbf{+0.96}$	$\mathbf{+0.96}$	$\mathbf{+0.96}$
	High	$+0.64$	$\mathbf{+1.00}$	$\mathbf{+0.89}$	$\mathbf{+0.71}$	$\mathbf{+0.96}$	$+0.36$

Table B.6: Phase 4 scalar load sensitivity, full-flushed grease (groups 3-1 and 3-2).

B.5. Waveform-Level Load-Sensitivity Matrices

The waveform-level analysis recomputes the scalar features from the stored transients and adds the spectral shape descriptors, giving an 11×3 matrix per test. The cross-test summary over all eighteen tests is the waveform-derived block of table 4.3, and figure 4.2 annotates that summary with the spread s_r and the sign-consistency fraction ϕ. The figure carries strictly more information than the individual matrices, since it reports dispersion as well as mean, so the per-test matrices are not reproduced in full.

Figure B.14 gives the single triplet the chapter reads individually, the 3-2 full-flushed traveller node of section 4.2.4. Stroke speed increases from left to right, and cell colour gives the Spearman correlation of each waveform-derived quantity with F_z per frequency band. Two behaviours are visible. Low-band load coupling sharpens as stroke speed increases, while the mid and high bands are already load coupled at every speed. The $40\,\mathrm{kHz}$ to $100\,\mathrm{kHz}$ and $100\,\mathrm{kHz}$ to $300\,\mathrm{kHz}$ partial powers also move in opposite directions with load.



(a) 3-2-1, $0.095\,\mathrm{m\,s^{-1}}$ **(b)** 3-2-2, $0.140\,\mathrm{m\,s^{-1}}$ **(c)** 3-2-3, $0.165\,\mathrm{m\,s^{-1}}$

Figure B.14: Waveform-level load sensitivity, 3-2 speed triplet.

B.6. Sensitivity to the Retained Hit Population

The waveform-level analysis was repeated on three retained hit populations: all threshold-detected hits, $\mathrm{SNR} \geq 5$ and $\mathrm{SNR} \geq 10$. The screening definition and its threshold base are given per dataset in table 4.2. As set out in section 4.2.4, the screen changes the measurement population rather than the bearing operating state, so a feature whose correlation changes under screening is recorded as sensitive to the retained population and not as operationally unstable.

Tables B.7 and B.8 give the cross-test summaries under the two screened populations, for direct comparison with the unfiltered waveform-derived block of table 4.3. In both tables, values are mean Spearman correlations with F_z and parentheses give the number of load-coupled tests ($|r_s| \geq 0.7$). Means are taken over eighteen tests, except low-band peak frequency, which uses seventeen tests at $\mathrm{SNR} \geq 5$ and fifteen at $\mathrm{SNR} \geq 10$.

Read together, the three tables support the population-sensitivity verdicts of section 4.2.4, marked by † in table 4.4. Waveform RMS and waveform peak hold their direction and approximate strength across all three populations, hit rate strengthens in the low and mid bands and remains strong in the high band, and the shape and spectral descriptors weaken or reverse. The per-triplet matrices underlying these means are not reproduced, since the verdicts rest on cross-test behaviour rather than on any individual

test.

Quantity	Low band	Mid band	High band
Hit-rate	$+\mathbf{0.85}$ (15)	$+\mathbf{0.89}$ (16)	$+\mathbf{0.89}$ (16)
Waveform RMS	$+0.69$ (16)	$+\mathbf{0.76}$ (15)	$+0.65$ (13)
Waveform peak	$+\mathbf{0.71}$ (15)	$+\mathbf{0.75}$ (15)	$+0.66$ (13)
Crest factor	-0.60 (8)	-0.17 (4)	$+0.50$ (9)
Kurtosis	-0.47 (6)	-0.28 (3)	$+0.18$ (12)
Peak frequency	$+0.12$ (6)	-0.04 (6)	-0.38 (14)
Spectral centroid	-0.00 (9)	-0.32 (10)	-0.39 (18)
Spectral roll-off $95\,\%$	$+0.20$ (8)	-0.01 (3)	$+0.09$ (12)
Partial power $40\,\mathrm{kHz}$ to $100\,\mathrm{kHz}$	-0.09 (8)	$+0.58$ (12)	-0.09 (12)
Partial power $100\,\mathrm{kHz}$ to $300\,\mathrm{kHz}$	$+0.19$ (10)	-0.57 (11)	$+0.38$ (16)
Partial power $300\,\mathrm{kHz}$ to $1000\,\mathrm{kHz}$	$+0.53$ (5)	-0.40 (9)	-0.38 (16)

Table B.7: Waveform-level load sensitivity at $\mathrm{SNR} \geq 5$.

Quantity	Low band	Mid band	High band
Hit-rate	$+\mathbf{0.83}$ (15)	$+\mathbf{0.87}$ (16)	$+\mathbf{0.88}$ (16)
Waveform RMS	$+\mathbf{0.70}$ (15)	$+\mathbf{0.75}$ (15)	$+0.67$ (14)
Waveform peak	$+\mathbf{0.72}$ (15)	$+\mathbf{0.75}$ (15)	$+0.67$ (15)
Crest factor	-0.56 (7)	-0.20 (3)	$+0.50$ (8)
Kurtosis	-0.40 (3)	-0.41 (3)	$+0.17$ (12)
Peak frequency	-0.05 (5)	-0.06 (7)	-0.39 (14)
Spectral centroid	-0.04 (7)	-0.41 (9)	-0.40 (18)
Spectral roll-off $95\,\%$	$+0.09$ (5)	-0.04 (2)	$+0.15$ (9)
Partial power $40\,\mathrm{kHz}$ to $100\,\mathrm{kHz}$	$+0.12$ (3)	$+0.62$ (10)	-0.09 (12)
Partial power $100\,\mathrm{kHz}$ to $300\,\mathrm{kHz}$	$+0.06$ (5)	-0.62 (10)	$+0.37$ (16)
Partial power $300\,\mathrm{kHz}$ to $1000\,\mathrm{kHz}$	$+0.54$ (6)	-0.48 (8)	-0.37 (16)

Table B.8: Waveform-level load sensitivity at $\mathrm{SNR} \geq 10$.

B.7. TestBlock 2 Withdrawn Load Matrices

The TestBlock 2 load axis is withdrawn in section 4.3.3, because the detection threshold on the two channels carrying $99.16\,\%$ of the block was raised in step with the load sequence (table 4.6), so the two effects cannot be separated. The matrices in tables B.9 and B.10 are retained only as a record of the fault. They are neither evidence for nor against a load effect and are not carried into any assessment. For that reason their cells are not emboldened, unlike every other correlation table in this appendix: bold marks a candidate for interpretation, and no value here is one.

Both tables use stroke speed as the repeat unit, with values given as mean Spearman correlations across the four load recordings and parentheses giving the number of speed repeats reaching $|r_s| \geq 0.7$.

Table B.9 is the evidence referred to in section 4.3.3, where every apparently load-coupled quantity moves in the direction a rising threshold would produce on its own. In the low band, which holds $99.82\,\%$ of the retained activity (section 4.3.2), median amplitude reads $+0.93$ and both RMS measures $+1.00$, the expected consequence of truncating the amplitude distribution from below. Duration reads -0.80 and counts -0.67, the expected consequence of losing the low-amplitude leading and trailing portions of each transient. The mid- and high-band columns rest on the sparse populations of section B.3.2 and are reported for completeness only.

Table B.10 covers the same repeats under the same withdrawal. Waveform RMS and waveform peak both read $+1.00$ in the low band, matching the acquisition RMS of the scalar table. Under a rising detection threshold this agreement carries no information about load.

Quantity	Low band	Mid band	High band
Hit-rate	-0.13 (2/3)	$+0.67$ (2/3)	$+0.33$ (3/3)
RMS (acquisition)	$+1.00$ (3/3)	$+0.53$ (1/3)	$+0.27$ (3/3)
Energy	$+0.53$ (2/3)	-0.13 (0/3)	$+0.60$ (2/3)
Amplitude [dB]	$+0.93$ (3/3)	-0.40 (1/3)	$+0.53$ (2/3)
Counts	-0.67 (2/3)	-0.27 (1/3)	$+0.60$ (2/3)
Duration	-0.80 (3/3)	$+0.33$ (0/3)	$+0.60$ (2/3)
Rise time	-0.80 (3/3)	$+0.20$ (0/3)	$+0.67$ (2/3)
Duration / counts	$+0.27$ (1/3)	$+0.13$ (0/3)	-0.40 (1/3)
Energy / duration	$+1.00$ (3/3)	-0.40 (1/3)	$+0.40$ (1/3)

Table B.9: TestBlock 2 scalar load correlations (withdrawn).

Quantity	Low band	Mid band	High band
Hit-rate	-0.13 (2/3)	$+0.67$ (2/3)	$+0.33$ (3/3)
Waveform RMS	$+1.00$ (3/3)	-0.20 (0/3)	$+0.47$ (1/3)
Waveform peak	$+1.00$ (3/3)	-0.40 (1/3)	$+0.53$ (2/3)
Crest factor	-0.80 (3/3)	$+0.00$ (0/3)	-0.13 (0/3)
Kurtosis	-0.67 (2/3)	-0.13 (0/3)	-0.33 (1/3)
Peak frequency	-0.16 (1/3)	-0.47 (1/3)	$+0.27$ (0/3)
Spectral centroid	$+0.53$ (2/3)	-0.47 (1/3)	-0.07 (1/3)
Spectral roll-off $95\,\%$	$+0.30$ (0/3)	-0.13 (0/3)	$+0.07$ (0/3)
Partial power $40\,\mathrm{kHz}$ to $100\,\mathrm{kHz}$	$+0.67$ (2/3)	$+0.27$ (0/3)	$+0.53$ (2/3)
Partial power $100\,\mathrm{kHz}$ to $300\,\mathrm{kHz}$	$+0.33$ (0/3)	-0.40 (1/3)	$+0.00$ (0/3)
Partial power $300\,\mathrm{kHz}$ to $1000\,\mathrm{kHz}$	-0.13 (1/3)	$+0.13$ (0/3)	$+0.00$ (0/3)

Table B.10: TestBlock 2 waveform-level load correlations (withdrawn).

B.8. TestBlock 3 Recording Map

The fourteen TestBlock 3 contamination files do not correspond one to one with the sixteen (run, load) combinations of the experimental programme. Several files combine two or three load steps, and two load steps are split across a file boundary. Load and speed are therefore reconstructed per hit from the operator markers rather than read from the file list. A separate pulsing calibration recording is excluded throughout.

Table B.11 gives the reconstructed mapping, with file names abbreviated from `20201015_contamination_`n. The speed- profile phase numbers the five legs of the symmetric ramp in order: position 1 is Speed 0 ascending, position 2 Speed 1 ascending, position 3 the Speed 2 apex, position 4 Speed 1 descending and position 5 Speed 0 descending. A complete pass spans positions 1–5, and any shorter span is a protocol exception.

Three protocol exceptions follow from the mapping and are carried through the analysis as stated here.

- **Baseline Load 3 is ascending only.** The pass covers speed- profile phases 1–3 and stops at the apex, so Speed 1 and Speed 0 are visited once rather than twice. It cannot contribute to an ascending-versus-descending comparison of the kind performed for TestBlock 2 in section 4.3.5.
- **Sample 1 Load 1 was run twice.** The second pass covers positions 1–3 with a truncated start. It is retained but flagged as non-primary, and is the source of the nineteen non-primary windows held back in chapter 6.
- **Sample 3 Loads 0 and 1 are each split across two files.** In both cases the break falls within the Speed 2 apex leg, which is rejoined during reconstruction. No hits are lost at the boundary, but the apex leg of these two passes is the only place in the block where a single speed segment spans two files.

All other passes follow the standard five-leg protocol. Passes recording no hits are treated as acquisition faults and excluded rather than entered as zero activity, since a zero would otherwise be read as a measured absence of emission.

Acquisition file	Run	Load step	Speed phase	Note
`.._contamination_1`	Baseline	0 ($100\,\mathrm{kN}$)	1–5	
`.._contamination_2`	Baseline	1 ($260\,\mathrm{kN}$)	1–5	
`.._contamination_3`	Baseline	2 ($785\,\mathrm{kN}$)	1–5	
`.._contamination_3`	Baseline	3 ($1305\,\mathrm{kN}$)	1–3	Ascending legs only; pass protocol incomplete
`.._contamination_4`	Sample 1	0 ($100\,\mathrm{kN}$)	1–5	
`.._contamination_5`	Sample 1	1 ($260\,\mathrm{kN}$)	1–5	
`.._contamination_6`	Sample 1	1 ($260\,\mathrm{kN}$)	1–3	Additional repeat, start truncated; non-primary
`.._contamination_6`	Sample 1	2 ($785\,\mathrm{kN}$)	1–5	
`.._contamination_6`	Sample 1	3 ($1305\,\mathrm{kN}$)	1–5	
`.._contamination_7`	Sample 2	0 ($100\,\mathrm{kN}$)	1–5	
`.._contamination_8`	Sample 2	1 ($260\,\mathrm{kN}$)	1–5	
`.._contamination_9`	Sample 2	2 ($785\,\mathrm{kN}$)	1–5	
`.._contamination_9`	Sample 2	3 ($1305\,\mathrm{kN}$)	1–5	
`.._contamination_10`	Sample 3	0 ($100\,\mathrm{kN}$)	1–3	Pass split at the apex
`.._contamination_11`	Sample 3	0 ($100\,\mathrm{kN}$)	3–5	Continuation of the above
`.._contamination_12`	Sample 3	1 ($260\,\mathrm{kN}$)	1–3	Pass split at the apex
`.._contamination_13`	Sample 3	1 ($260\,\mathrm{kN}$)	3–5	Continuation of the above
`.._contamination_14`	Sample 3	2 ($785\,\mathrm{kN}$)	1–5	
`.._contamination_14`	Sample 3	3 ($1305\,\mathrm{kN}$)	1–5	

Table B.11: TestBlock 3 file to operating-state mapping.

B.9. Speed-Sensitivity Summary

This section supports the Phase 4 speed discussion of section 4.2.5. The speed response was evaluated within each of the six grouped test conditions, formed by lubrication condition $\times$ sensor node position, by correlating each feature against stroke speed and recording the largest relative change across the three tested speeds. Each condition was then classified as a strong speed trend, a speed-related but non-monotonic response, or a weak or absent speed effect. Because these comparisons cross recordings, and therefore sensor re-couplings and ambient conditions, they are read as a distribution rather than as six independent estimates. The chapter's constant-load evidence, which is free of that confound, is figure 4.3.

Table B.12 gives the distribution. For each feature and band, the first three columns give the number of conditions in each class, summing to six in every row, and the final two columns give the median and maximum of the largest relative change across the six conditions.

The table supports the three claims made in section 4.2.5. Median amplitude is the most speed-stable quantity in the set, weak in all six conditions in the low band and in five of six in the high band, with median largest changes of $2.5\,\%$ and $5.8\,\%$ respectively. Hit rate is the least stable, with a median largest change of $55\,\%$ to $114\,\%$ depending on band and no condition classified as weak in any band. Non-monotonic responses are common rather than exceptional: fifty of the 162 feature–band–condition combinations are so classified, which is why stroke speed is treated as an operational modifier rather than fitted as a trend across recordings.

		Test conditions (of 6)			Largest change [%]	
Feature	Band	Strong	Non-mono.	Weak	Median	Max.
`hit_rate`	low	2	4	0	55.5	378.1
	mid	2	4	0	69.8	518.9
	high	3	3	0	114.0	635.1
`Amp_dB`	low	0	0	6	2.5	3.7
	mid	2	1	3	7.8	17.4
	high	0	1	5	5.8	11.0
`RMS`	low	5	1	0	38.4	43.1
	mid	2	3	1	14.8	26.7
	high	5	0	1	21.1	26.3
`Eny`	low	4	2	0	30.4	61.3
	mid	4	2	0	93.4	639.7
	high	4	1	1	53.2	86.2
`Counts`	low	3	1	2	24.6	36.0
	mid	6	0	0	64.4	239.2
	high	1	2	3	10.4	50.8
`Dur`	low	1	3	2	11.2	23.4
	mid	5	1	0	48.2	64.2
	high	0	2	4	3.8	34.5
`RiseT`	low	4	1	1	23.2	32.1
	mid	3	3	0	32.3	58.9
	high	4	2	0	24.3	68.5
`Dur_per_Count`	low	4	1	1	16.5	25.5
	mid	2	4	0	26.4	74.0
	high	0	1	5	2.8	12.2
`Eny_per_Dur`	low	4	1	1	29.4	45.9
	mid	2	4	0	54.9	387.4
	high	4	2	0	53.3	77.8

Table B.12: Phase 4 speed-effect summary across the six test conditions.

B.10. Lubrication-Sensitivity Table

This section supports the Phase 4 lubrication discussion of section 4.2.6. In table B.13 the mean ratio compares the feature level under each lubrication condition with the corresponding lifetime-grease reference at matched speeds. Values above one indicate higher AE levels than lifetime grease and values below one lower levels. The percentage column expresses the same difference as a relative change, and the final column classifies the difference under the analysis threshold.

Because film thickness and grease distribution were not measured, these values describe the imposed grease state rather than a measured lubrication regime. They are reported as descriptive evidence for the band and feature dependence asserted in section 4.2.6, and lubrication is not treated as a correction axis anywhere in this thesis.

Table B.13: Phase 4 lubrication effect relative to lifetime grease.

Feature	Band	Mean ratio to lifetime	Mean change vs. lifetime (%)	n speeds	Lubrication-effect flag
Full mixed					
Amp_dB	high	1.072	7.2	3	small/no lubrication difference
Amp_dB	low	1.007	0.7	3	small/no lubrication difference
Amp_dB	mid	1.093	9.3	3	small/no lubrication difference
Counts	high	1.194	19.4	3	large lubrication difference
Counts	low	1.119	11.9	3	small/no lubrication difference
Counts	mid	2.205	120.5	3	large lubrication difference
Dur	high	1.105	10.5	3	large lubrication difference
Dur	low	1.072	7.2	3	small/no lubrication difference
Dur	mid	1.510	51.0	3	large lubrication difference
Dur_per_Count	high	0.948	-5.2	3	small/no lubrication difference
Dur_per_Count	low	0.941	-5.9	3	small/no lubrication difference
Dur_per_Count	mid	0.645	-35.5	3	large lubrication difference
Eny	high	2.314	131.4	3	large lubrication difference
Eny	low	1.176	17.6	3	large lubrication difference
Eny	mid	4.090	309.0	3	large lubrication difference
Eny_per_Dur	high	2.049	104.9	3	large lubrication difference
Eny_per_Dur	low	1.073	7.3	3	small/no lubrication difference
Eny_per_Dur	mid	2.541	154.1	3	large lubrication difference
RMS	high	1.018	1.8	3	small/no lubrication difference
RMS	low	0.961	-3.9	3	small/no lubrication difference
RMS	mid	1.013	1.3	3	small/no lubrication difference
RiseT	high	0.807	-19.3	3	large lubrication difference
RiseT	low	1.054	5.4	3	small/no lubrication difference
RiseT	mid	1.132	13.2	3	large lubrication difference
hit_rate	high	1.438	43.8	3	large lubrication difference
hit_rate	low	1.045	4.5	3	large lubrication difference
hit_rate	mid	1.246	24.6	3	large lubrication difference
Full flushed					
Amp_dB	high	1.022	2.2	3	small/no lubrication difference
Amp_dB	low	0.996	-0.4	3	small/no lubrication difference
Amp_dB	mid	1.022	2.2	3	small/no lubrication difference
Counts	high	0.818	-18.2	3	large lubrication difference
Counts	low	1.119	11.9	3	large lubrication difference
Counts	mid	1.498	49.8	3	large lubrication difference
Dur	high	0.851	-14.9	3	large lubrication difference
Dur	low	1.116	11.6	3	large lubrication difference
Dur	mid	1.194	19.4	3	large lubrication difference
Dur_per_Count	high	1.034	3.4	3	small/no lubrication difference
Dur_per_Count	low	0.991	-0.9	3	small/no lubrication difference
Dur_per_Count	mid	0.901	-9.9	3	large lubrication difference
Eny	high	0.802	-19.8	3	large lubrication difference
Eny	low	1.080	8.0	3	large lubrication difference
Eny	mid	2.931	193.1	3	large lubrication difference
Eny_per_Dur	high	1.015	1.5	3	large lubrication difference
Eny_per_Dur	low	0.975	-2.5	3	small/no lubrication difference
Eny_per_Dur	mid	1.380	38.0	3	large lubrication difference

Continued on next page

Feature	Band	Mean ratio to lifetime	Mean change vs. lifetime (%)	n speeds	Lubrication-effect flag
RMS	high	1.093	9.3	3	small/no lubrication difference
RMS	low	1.070	7.0	3	large lubrication difference
RMS	mid	1.048	4.8	3	large lubrication difference
RiseT	high	0.642	-35.8	3	large lubrication difference
RiseT	low	1.117	11.7	3	large lubrication difference
RiseT	mid	0.961	-3.9	3	large lubrication difference
hit_rate	high	1.465	46.5	3	large lubrication difference
hit_rate	low	1.127	12.7	3	large lubrication difference
hit_rate	mid	1.254	25.4	3	large lubrication difference

C

Full-Scale Swivel Bearing: Supporting Data

This appendix collects the supporting material for the full-scale swivel-bearing experiment of Chapter 5: the complete test matrix, the bearing dimensions, the uncalibrated pressure traces, the operator rotation logs, the preprocessing pipeline, the hit activity through each recording, the amplitude-screening sensitivity, the individual signed correlation estimates for the complete pressure sweep, and the directional repeatability results.

C.1. Complete BTF Test Matrix

Table C.1 summarises all recordings obtained during the BTF campaign. The six recordings from SB-L2-SR to SB-L7-SR form the complete same-day pressure sweep used in the operational-influence analysis; the remaining recordings are retained for later repeatability and data-quality assessment, and the final column states why each falls outside the selected sweep.

Table C.1: Complete BTF test programme and dataset selection.

Test case	Pressure (barg)	Date	Duration (min)	Total hits	Active channels	Median amp. (dB)	Exclusion notes
SB-L0-SR	0	2026-07-07	44.5	412	9	54.8	Preliminary recording acquired before the selected same-day sweep.
SB-L1-SR-A	20	2026-07-07	80.8	2259	7	50.8	First-day recording outside the selected sweep; two channels inactive.
SB-L1-SR-B	20	2026-07-07	44.0	1193	9	53.2	First-day repeat recording outside the selected sweep.
SB-L2-SR	25	2026-07-08	45.0	609	9	52.8	
SB-L3-SR	20	2026-07-08	44.1	11576	9	51.6	
SB-L4-SR	15	2026-07-08	45.9	76388	9	52.8	
SB-L5-SR	10	2026-07-08	42.8	77311	9	53.4	
SB-L6-SR	5	2026-07-08	43.2	15797	9	53.3	
SB-L7-SR	0	2026-07-08	44.4	3332	9	53.2	
SB-L8-SR	0	2026-07-08	17.9	54	3	50.0	Short post-sweep recording with six inactive channels and limited yield.

C.2. Bearing Dimensions

The main dimensions of the BTF swivel bearing are listed in table C.2. The outside diameter, bore diameter and overall height are documented values from the available manufacturer drawing. Additional internal dimensions were obtained from the drawing where possible and are treated as approximate, since no further dimensional documentation was available from Bluewater Energy Services or the bearing supplier. The basis for each entry is stated in the final column of the table.

The raceway pitch-circle diameter D_{pcd}, required to convert angular speed to raceway surface speed, is not specified on the drawing. It was estimated from the section view using the dimensioned bearing geometry, giving

$$D_{\mathrm{pcd}} = 1163\,\mathrm{mm}. \tag{C.1}$$

The radial-raceway roller diameter and length were measured from the drawing, and the number of radial rollers was estimated from the section view; all three are used as engineering estimates rather than documented values.

Feature	Symbol	Value	Basis
Outer ring outside diameter	D_{o}	$1315\,\mathrm{mm}$	Documented
Outer ring spigot diameter	–	$1313\,\mathrm{mm}$	Drawing estimate
Outer seal / shoulder diameter	–	$1288\,\mathrm{mm}$	Drawing estimate
Inner ring outside diameter	–	$1072\,\mathrm{mm}$	Drawing estimate
Inner ring register diameter	–	$1027\,\mathrm{mm}$	Drawing estimate
Bore diameter	D_{i}	$1018\,\mathrm{mm}$	Documented
Outer seal groove diameter	D_{seal}	$962\,\mathrm{mm}$	Drawing estimate
Raceway pitch-circle diameter	D_{pcd}	$1163\,\mathrm{mm}$	Geometry estimate
Overall height	H	$130\,\mathrm{mm}$	Documented
Outer ring height	–	$120\,\mathrm{mm}$	Drawing estimate
Inner ring height	–	$73\,\mathrm{mm}$	Drawing estimate
Roller diameter, radial raceway	d_{r}	$16\,\mathrm{mm}$	Drawing estimate
Roller length, radial raceway	l_{r}	$16\,\mathrm{mm}$	Drawing estimate
Number of rollers, radial raceway	Z	184	Engineering estimate

Table C.2: Documented and estimated dimensions of the BTF swivel bearing.

C.3. Uncalibrated Pressure Traces

The pressure traces for the two acquisition days are given in figures C.1 and C.2. They are uncalibrated and are reproduced to document the pressure history of each day, including the order and timing of the commanded steps, rather than to provide absolute pressure values.

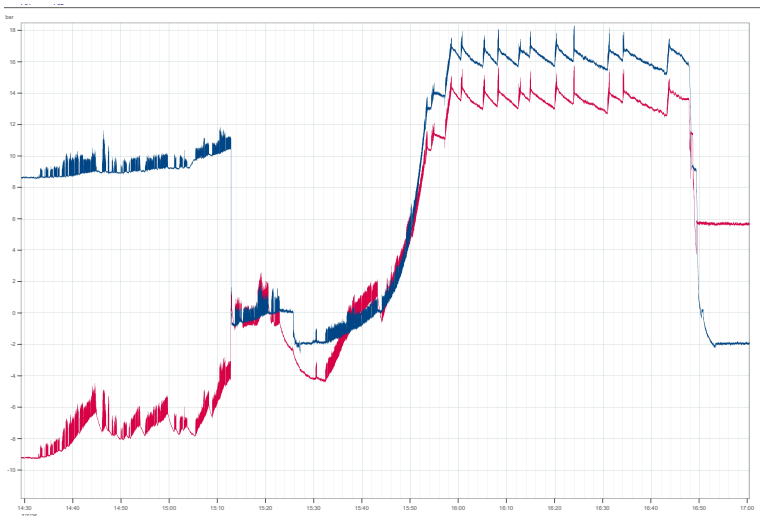


Figure C.1: Uncalibrated pressure trace, 7 July 2026.

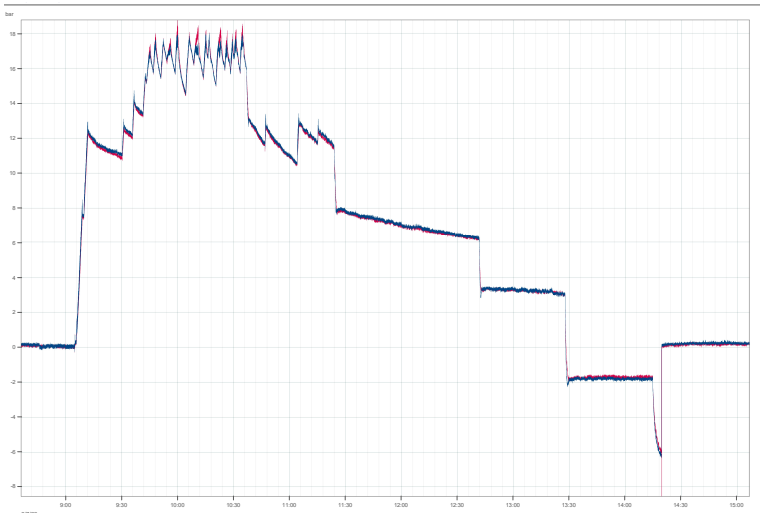


Figure C.2: Uncalibrated pressure trace, 8 July pressure sweep.

C.4. Operator Rotation Logs

The operator rotation logs used for the time alignment are listed in tables C.3 and C.4. Each entry gives the wall-clock start time of the corresponding commanded rotation pass. Positive rotation speeds denote anticlockwise motion and negative speeds denote clockwise motion.

	Test case	
Speed [deg/s]	SB-L0-SR (0 barg)	SB-L1-SR (20 barg)
0.5	13:48	16:02
-0.5	13:59	16:14
1.5	14:12	16:27
-1.5	14:16	16:31
3.0	14:21	16:36
-3.0	14:23	16:38
4.5	14:26	16:40
-4.5	14:28	16:42
6.0	14:30	16:44
-6.0	14:31	16:45

Table C.3: Operator rotation logs, 7 July 2026.

	Test case					
Speed [deg/s]	SB-L2-SR (25 barg)	SB-L3-SR (20 barg)	SB-L4-SR (15 barg)	SB-L5-SR (10 barg)	SB-L6-SR (5 barg)	SB-L7-SR (0 barg)
0.5	09:50	10:38	11:26	12:43	13:31	14:20
-0.5	10:03	10:51	11:40	12:55	13:43	14:33
1.5	10:15	11:03	11:52	13:08	13:55	14:45
-1.5	10:20	11:08	11:56	13:12	14:00	14:49
3.0	10:24	11:12	12:01	13:16	14:04	14:54
-3.0	10:27	11:14	12:03	13:19	14:06	14:56
4.5	10:30	11:17	12:05	13:21	14:08	14:59
-4.5	10:31	11:18	12:07	13:22	14:10	15:00
6.0	10:33	11:20	12:08	13:24	14:12	15:02
-6.0	10:34	11:21	12:09	13:25	14:13	15:03

Table C.4: Operator rotation logs, 8 July pressure sweep.

C.5. Preprocessing Pipeline

Table C.5 gives the implementation of the preprocessing stages of section 3.3 for the BTF dataset. Counts refer to the complete recorded population of $185\,013$ hits. Unlike the linear-bearing datasets, the BTF experiment has no continuous operational data stream, and the amplitude screen is an absolute recorded-amplitude gate rather than a measured peak-to-noise ratio.

Stage	BTF swivel bearing
Validation and units	Hit records on the nine analysis channels only; node N4 is excluded at source. Every record carries a valid channel, a positive amplitude and a valid transient index, so no record is removed. Amplitudes are converted to dB relative to $1\,\mu\mathrm{V}$; each channel is mapped to its sensor node and frequency band
Saturation and censoring	Neither mechanism is active. The largest recorded amplitude is $20\,999.9\,\mu\mathrm{V}$ against the ceiling $A_{\max} = 49\,990\,\mu\mathrm{V}$, and the longest event is $22\,475\,\mu\mathrm{s}$ against the $104\,857.6\,\mu\mathrm{s}$ ceiling (section 5.3.1)
Amplitude and SNR screening	An absolute gate on recorded hit amplitude: $55\,\mathrm{dB}$ in the low band and $50\,\mathrm{dB}$ in the mid and high bands, stored as the per hit flag `snr_pass` and applied non-destructively. Retains $135\,481$ of $185\,013$ hits ($73.2\,\%$), unevenly by band (table C.6)
Operational cleaning and synchronisation	No operational data stream exists. The stage reduces to referring the operator rotation log to the first AE hit of each recording (equation (5.3)), at minute resolution and subject to the lag established in section 5.7
Segmentation into analysis units	Pressure is the recording. Within it, one contiguous window per logged pass, $[\tau_{j,k},\ \tau_{j,k+1})$, giving ten windows per recording and sixty in total. The analysis unit is the pressure–speed–channel combination, 270 in total, with activity expressed both per minute and per nominal rotation

Table C.5: BTF preprocessing pipeline.

C.6. Hit Activity Through Each Recording

Figure C.3 shows the cumulative hit count through each recording of the selected sweep on a common logarithmic vertical scale. Dotted lines mark speed-window boundaries from the operator log, and commanded speed is shown above each panel. At 0–$15\ \mathrm{barg}$, activity occurs at every speed and band ordering is preserved; at 20 and $25\ \mathrm{barg}$, activity shifts strongly towards the two fastest levels.

C.7. Amplitude-Screening Sensitivity

At the adopted thresholds of section 5.3.1, the screen removes $27\,580$ low-band, $21\,264$ mid-band and 688 high-band hits, $49\,532$ in total ($26.8\,\%$). Table C.6 gives the retained fractions under the adopted and alternative screens. Table C.7 reports selected relationships over the complete sweep under each screen; entries are mean Spearman correlations, parentheses give coupling counts, and the adopted column is the setting used throughout Chapter 5.

Screen	Thresholds	Retained (%)			
	(low/mid/high, dB)	All	Low	Mid	High
None	—	100.0	100.0	100.0	100.0
Adopted	55/50/50	73.2	30.7	84.4	92.3
$+5\ \mathrm{dB}$	60/55/55	27.2	8.2	32.0	39.8
$+10\ \mathrm{dB}$	65/60/60	8.0	2.0	9.5	11.6

Table C.6: Retention under alternative amplitude screens.

C.8. Individual Signed Correlations

The sensitivity tables of Chapter 5 report the reduced 0–$15\ \mathrm{barg}$ sweep. This section gives the complete six-recording sweep as individual Spearman estimates: five for the pressure axis, one per commanded

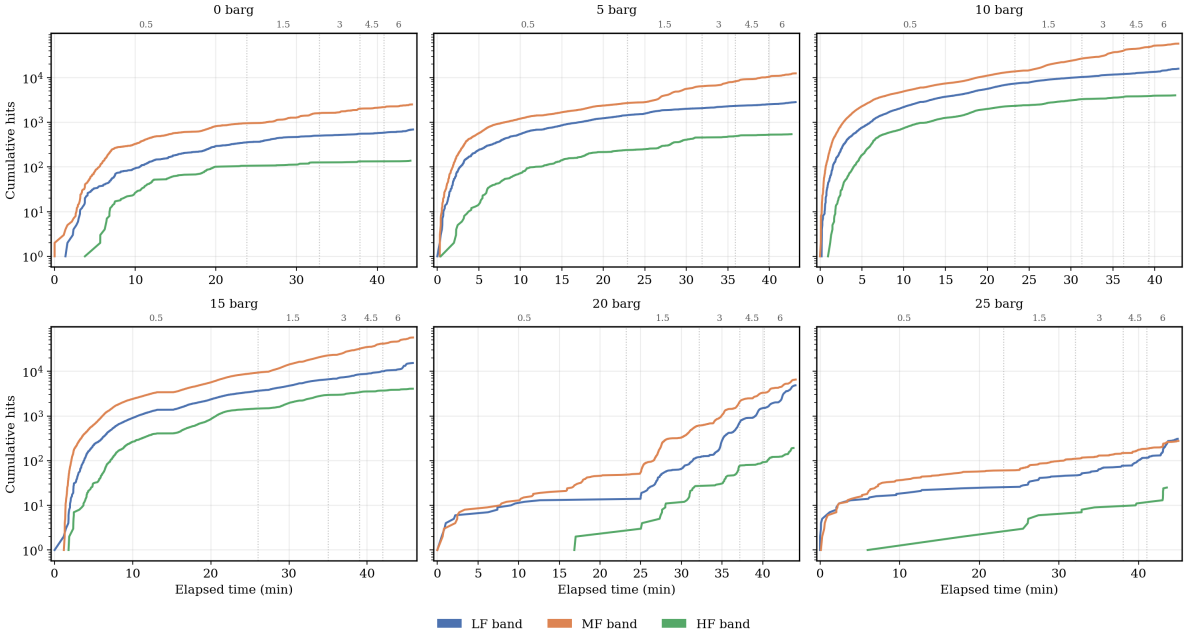


Figure C.3: Cumulative hits by pressure and frequency band.

Relationship	Band	None	Adopted	+5 dB	+10 dB
Hit rate vs. pressure	Mid	−0.28 (0/5)	−0.28 (0/5)	−0.29 (0/5)	−0.22 (0/5)
Crest factor vs. pressure	Mid	−0.54 (2/5)	**−0.77** (4/5)	**−0.75** (3/5)	−0.42 (1/5)
Kurtosis vs. pressure	Mid	−0.68 (3/5)	**−0.81** (4/5)	**−0.70** (4/5)	−0.47 (2/5)
Waveform RMS vs. speed	Mid	**−0.73** (5/6)	−0.53 (4/6)	−0.62 (4/6)	−0.67 (6/6)
Waveform peak vs. speed	Low	**−0.92** (6/6)	−0.65 (4/6)	−0.45 (3/6)	−0.20 (2/5)

Table C.7: Sensitivity of selected relationships to amplitude screening.

speed level, and six for the speed axis, one per pressure recording. The Mean column is the complete-sweep mean quoted in the chapter. Where estimates of opposite sign are averaged, the mean is not a meaningful summary; these tables allow this to be checked directly for any quantity rather than inferred from the coupling count. For the speed axis, the reduced-sweep values in table 5.3 correspond to the first four columns (0–$15\ \mathrm{barg}$). The reduced-sweep pressure estimates are recomputed over four recordings and are not tabulated here.

In the two pressure-sensitivity tables each column is one commanded speed level in $\mathrm{deg\,s^{-1}}$, and each entry is the Spearman correlation with gauge pressure across the six recordings at that speed. In the two speed-sensitivity tables each column is one pressure recording in barg, and each entry is the correlation with commanded speed across the five speed levels of that recording. Estimates in bold clear the association screen $|r_s| \geq 0.7$ of section 3.5, and a bullet in the final column marks a quantity–band row whose estimates are not all of one sign. All populations are the screened hits of section 5.3.1.

Such rows are the majority in every table: 20 of 24 for the hit features against pressure, 22 of 30 for the waveform features against pressure, and 20 of 24 and 24 of 30 for the same two sets against speed. The mean spread between the largest and smallest estimate behind a single reported value is 0.98 and 0.92 on the pressure axis and 1.56 and 1.32 on the speed axis. The speed axis is therefore the less stable of the two: a spread of 1.56 on a coefficient bounded by ± 1 means that the six recordings frequently disagree from one extreme to the other. Mid-band energy/duration, for example, has a complete-sweep mean of -0.67 but reaches -1.00 in four of the six recordings and $+0.70$ in the sixth. The speed means should therefore be read with at least the caution applied to the pressure means.

The reversals concentrate at the edges of each axis. On the pressure axis, the slowest speed level disagrees with the other four more often than any other column. On the speed axis, the 20 and $25\ \mathrm{barg}$ recordings supply most of the reversals, consistent with the sparse and speed-confounded populations described in section 5.7 and a further reason for reporting the reduced sweep in the main text. The quantities that carry the chapter's conclusions behave differently: low-band amplitude against speed, mid-band crest factor and mid-band kurtosis against pressure, and low-band waveform peak against speed are among the few rows with no reversal.

Hit rate appears here in its per-minute form only. Complete-sweep means for the per-rotation variant against speed are quoted in section 5.4.2.

C.9. Directional Repeatability

Table C.12 reports the median symmetric relative difference between anticlockwise and clockwise visits at the same pressure and speed, using 24, 49 and 16 paired pressure–speed–channel combinations in the low, mid and high bands (section 5.5). The verdict reports the worst band.

Quantity	0.5	1.5	3	4.5	6	Mean	Rev.
Low band							
Hit rate	-0.43	-0.49	+0.26	+0.26	+0.37	-0.01	•
Duration	-0.09	+0.26	**+0.71**	**+0.89**	+0.37	+0.43	•
Energy	-0.60	+0.03	+0.09	**+0.71**	+0.37	+0.12	•
Amplitude (dB)	-0.54	+0.03	+0.37	-0.09	-0.09	-0.06	•
Rise time	**-0.71**	**+0.83**	+0.60	+0.37	+0.09	+0.23	•
Counts	+0.09	+0.03	**+0.84**	**+0.93**	+0.35	+0.45	
Duration / counts	-0.60	-0.03	**-0.71**	-0.26	+0.09	-0.30	•
Energy / duration	-0.03	+0.43	-0.09	+0.03	-0.03	+0.06	•
Mid band							
Hit rate	-0.43	-0.49	-0.26	-0.20	-0.03	-0.28	
Duration	-0.54	+0.60	**+0.77**	**+0.71**	**+0.83**	+0.47	•
Energy	-0.43	+0.60	+0.31	+0.60	**+0.83**	+0.38	•
Amplitude (dB)	-0.49	+0.14	-0.37	**-0.77**	+0.03	-0.29	•
Rise time	+0.26	**+0.77**	+0.60	**+0.94**	**+0.83**	+0.68	
Counts	-0.43	+0.60	+0.44	**+0.81**	**+0.90**	+0.46	•
Duration / counts	-0.09	-0.43	**+0.77**	**+0.71**	-0.20	+0.15	•
Energy / duration	-0.09	-0.20	-0.60	**-0.77**	+0.03	-0.33	•
High band							
Hit rate	-0.49	-0.20	-0.03	+0.09	+0.26	-0.07	•
Duration	+0.37	+0.26	+0.09	**+0.70**	+0.09	+0.30	
Energy	-0.43	**+0.83**	+0.03	+0.60	+0.20	+0.25	•
Amplitude (dB)	-0.09	+0.03	+0.09	+0.30	+0.66	+0.20	•
Rise time	+0.26	+0.26	-0.37	+0.10	**-0.77**	-0.11	•
Counts	+0.20	+0.49	+0.09	+0.36	-0.03	+0.22	•
Duration / counts	-0.14	+0.54	**-0.71**	-0.40	-0.43	-0.23	•
Energy / duration	-0.26	**-1.00**	+0.03	+0.40	+0.49	-0.07	•

Table C.8: Individual pressure-sensitivity estimates, hit features, complete sweep.

Quantity	0.5	1.5	3	4.5	6	Mean	Rev.
Low band							
RMS	**-0.77**	+0.03	-0.03	+0.31	+0.43	-0.01	•
Peak amplitude	-0.54	+0.03	+0.37	-0.09	-0.09	-0.06	•
Crest factor	+0.20	+0.20	+0.43	**+0.77**	+0.26	+0.37	
Kurtosis	-0.37	+0.66	+0.49	**-0.71**	+0.26	+0.06	•
Peak frequency	+0.03	-0.14	+0.26	+0.23	+0.26	+0.13	•
Spectral centroid	+0.03	+0.03	+0.37	+0.49	-0.09	+0.17	•
Spectral roll-off	+0.26	+0.20	+0.37	+0.49	+0.43	+0.35	
LF partial power	+0.43	-0.09	+0.54	+0.26	-0.09	+0.21	•
MF partial power	+0.26	+0.37	+0.66	**+0.77**	+0.49	+0.51	
HF partial power	-0.03	**-0.83**	**-0.83**	-0.49	+0.60	-0.31	•
Mid band							
RMS	-0.43	+0.60	+0.14	+0.60	**+0.77**	+0.34	•
Peak amplitude	-0.49	+0.14	-0.37	**-0.77**	+0.03	-0.29	•
Crest factor	**-0.83**	-0.43	**-0.94**	**-0.89**	**-0.77**	-0.77	
Kurtosis	**-0.77**	-0.49	**-1.00**	**-1.00**	**-0.77**	-0.81	
Peak frequency	**+0.77**	+0.43	+0.54	+0.37	+0.60	+0.54	
Spectral centroid	**+0.77**	-0.31	-0.14	-0.03	-0.09	+0.04	•
Spectral roll-off	+0.20	-0.03	-0.03	-0.20	-0.09	-0.03	•
LF partial power	+0.49	+0.31	+0.37	+0.09	+0.26	+0.30	
MF partial power	-0.66	-0.31	**-0.77**	-0.03	**-0.83**	-0.52	
HF partial power	**+0.77**	-0.03	**+0.77**	**+0.83**	+0.60	+0.59	•
High band							
RMS	-0.49	+0.26	+0.03	+0.40	**+0.77**	+0.19	•
Peak amplitude	-0.09	+0.03	+0.09	+0.30	+0.66	+0.20	•
Crest factor	+0.14	+0.37	+0.14	-0.20	+0.60	+0.21	•
Kurtosis	+0.09	**-0.89**	-0.66	-0.10	**+0.94**	-0.12	•
Peak frequency	+0.31	-0.20	**+1.00**	**+0.90**	-0.09	+0.39	•
Spectral centroid	**+0.89**	**+0.94**	+0.66	**+0.90**	-0.54	+0.57	•
Spectral roll-off	**+0.94**	**+0.83**	**+0.77**	**+1.00**	-0.14	+0.68	•
LF partial power	+0.43	**-1.00**	+0.09	+0.10	+0.43	+0.01	•
MF partial power	**-0.77**	**+0.83**	+0.54	-0.60	+0.31	+0.06	•
HF partial power	+0.37	**-0.83**	-0.54	-0.20	-0.31	-0.30	•

Table C.9: Individual pressure-sensitivity estimates, waveform features, complete sweep.

Quantity	0	5	10	15	20	25	Mean	Rev.
Low band								
Hit rate	+0.10	**-0.90**	**-0.90**	+0.20	**+1.00**	**+1.00**	+0.08	•
Duration	+0.10	**-0.70**	-0.60	+0.40	**+1.00**	+0.10	+0.05	•
Energy	+0.30	**-1.00**	**-0.90**	-0.40	+0.60	-0.20	-0.27	•
Amplitude (dB)	**-0.80**	**-1.00**	**-1.00**	**-0.90**	-0.20	+0.00	-0.65	
Rise time	+0.30	**-0.80**	+0.10	**+0.70**	**+0.70**	+0.30	+0.22	•
Counts	+0.00	**-0.70**	**-1.00**	+0.22	**+1.00**	+0.10	-0.06	•
Duration / counts	**+0.80**	**+0.90**	**+1.00**	**+1.00**	**+0.90**	+0.20	+0.80	
Energy / duration	-0.50	-0.60	**-1.00**	**-1.00**	**-0.90**	+0.00	-0.67	
Mid band								
Hit rate	**+0.70**	**+1.00**	+0.60	**+1.00**	**+1.00**	**+1.00**	+0.88	
Duration	**-0.90**	**-0.70**	-0.60	-0.30	+0.60	+0.30	-0.27	•
Energy	**-0.90**	**-1.00**	**-1.00**	**-0.90**	+0.20	+0.60	-0.50	•
Amplitude (dB)	-0.40	**-0.90**	**-1.00**	**-1.00**	-0.50	+0.60	-0.53	•
Rise time	**-0.90**	**-0.90**	**-0.70**	+0.40	**+0.90**	+0.50	-0.12	•
Counts	**-0.87**	**-1.00**	**-1.00**	**-0.90**	+0.22	+0.67	-0.48	•
Duration / counts	**+0.70**	**+0.70**	**+1.00**	**+1.00**	**+1.00**	-0.30	+0.68	•
Energy / duration	**-0.70**	**-1.00**	**-1.00**	**-1.00**	**-1.00**	**+0.70**	-0.67	•
High band								
Hit rate	**-0.90**	**-0.80**	**-0.90**	+0.00	**+0.90**	**+0.90**	-0.13	•
Duration	**+0.80**	-0.40	**-0.90**	**-1.00**	+0.40	+0.60	-0.08	•
Energy	+0.00	**-0.90**	**-0.90**	**-1.00**	-0.50	**+0.90**	-0.40	•
Amplitude (dB)	-0.20	**-0.90**	**-1.00**	**-0.90**	+0.00	**+0.90**	-0.35	•
Rise time	**+0.80**	+0.30	-0.50	**-0.70**	+0.30	**-0.82**	-0.10	•
Counts	+0.60	-0.40	**-0.89**	**-1.00**	-0.20	**+0.70**	-0.20	•
Duration / counts	+0.40	-0.30	**+0.80**	**+1.00**	+0.50	**-0.90**	+0.25	•
Energy / duration	**-0.80**	+0.30	**-0.70**	**+0.70**	**-0.80**	**+1.00**	-0.05	•

Table C.10: Individual speed-sensitivity estimates, hit features, complete sweep.

Quantity	0	5	10	15	20	25	Mean	Rev.
Low band								
RMS	+0.20	**-0.70**	**-1.00**	**-0.90**	-0.50	+0.30	-0.43	•
Peak amplitude	**-0.80**	**-1.00**	**-1.00**	**-0.90**	-0.20	+0.00	-0.65	
Crest factor	-0.60	+0.20	+0.30	-0.40	**+0.80**	-0.60	-0.05	•
Kurtosis	-0.20	-0.60	+0.60	**-0.70**	**-0.90**	+0.10	-0.28	•
Peak frequency	+0.20	-0.20	+0.20	-0.15	+0.40	+0.30	+0.13	•
Spectral centroid	-0.60	**-0.70**	**-0.70**	**-0.90**	+0.00	-0.40	-0.55	
Spectral roll-off	**-0.70**	-0.60	**-0.90**	**-1.00**	-0.10	**-0.70**	-0.67	
LF partial power	+0.20	+0.10	**+0.70**	+0.60	-0.10	-0.50	+0.17	•
MF partial power	**-1.00**	**-0.80**	**-0.90**	**-0.90**	**+0.80**	-0.50	-0.55	•
HF partial power	**+0.70**	**+0.70**	**+0.90**	**+0.90**	+0.00	+0.40	+0.60	
Mid band								
RMS	**-0.80**	**-1.00**	**-1.00**	**-1.00**	+0.00	+0.60	-0.53	•
Peak amplitude	-0.56	**-0.90**	**-1.00**	**-1.00**	-0.56	+0.60	-0.57	•
Crest factor	-0.30	-0.60	-0.30	**-0.70**	-0.50	**+0.90**	-0.25	•
Kurtosis	-0.30	-0.10	**-1.00**	**-1.00**	**-1.00**	+0.60	-0.47	•
Peak frequency	**+0.70**	**+0.70**	**+0.70**	+0.60	-0.40	+0.10	+0.40	•
Spectral centroid	**-0.70**	-0.40	-0.30	-0.10	**-1.00**	-0.60	-0.52	
Spectral roll-off	-0.20	+0.05	+0.50	+0.60	+0.00	-0.15	+0.13	•
LF partial power	+0.60	**+0.70**	**+0.70**	**+1.00**	+0.00	**+1.00**	+0.67	
MF partial power	-0.60	**-1.00**	**-0.70**	**-0.70**	+0.10	**-0.80**	-0.62	•
HF partial power	-0.20	+0.40	**+1.00**	+0.60	-0.30	+0.20	+0.28	•
High band								
RMS	-0.40	**-1.00**	**-1.00**	**-0.90**	-0.40	**+1.00**	-0.45	•
Peak amplitude	-0.20	**-0.90**	**-0.90**	**-0.90**	-0.10	**+1.00**	-0.33	•
Crest factor	+0.60	-0.10	-0.20	**-0.90**	+0.00	+0.50	-0.02	•
Kurtosis	-0.20	-0.30	+0.60	+0.10	+0.00	**+0.90**	+0.18	•
Peak frequency	+0.40	-0.30	**-0.70**	**-1.00**	+0.50	+0.60	-0.08	•
Spectral centroid	**+1.00**	+0.30	-0.40	**-1.00**	**+0.80**	-0.50	+0.03	•
Spectral roll-off	**+0.80**	+0.30	+0.00	**-0.90**	+0.60	-0.20	+0.10	•
LF partial power	-0.40	-0.60	**+0.90**	**+0.90**	-0.10	+0.20	+0.15	•
MF partial power	-0.20	-0.60	+0.10	-0.40	+0.00	**+0.70**	-0.07	•
HF partial power	+0.20	+0.10	-0.10	+0.40	-0.50	-0.20	-0.02	•

Table C.11: Individual speed-sensitivity estimates, waveform features, complete sweep.

Quantity	Low band	Mid band	High band	Verdict
Amplitude (dB)	0.5	0.6	0.6	Repeatable
Peak frequency	0.5	0.2	1.1	Repeatable
Spectral roll-off	2.9	0.1	2.0	Repeatable
Spectral centroid	2.0	0.8	1.9	Repeatable
Crest factor	3.3	1.3	1.6	Repeatable
Waveform RMS	2.5	2.7	2.9	Repeatable
Waveform peak	3.1	3.6	3.8	Repeatable
Duration / counts	4.4	5.2	5.4	Repeatable
Energy / duration	6.9	6.7	9.1	Repeatable
Kurtosis	9.0	6.3	7.2	Repeatable
Energy	8.3	8.8	8.2	Repeatable
Counts	10.8	8.7	11.8	Marginal
Duration	11.3	10.4	11.9	Marginal
Rise time	14.6	11.7	8.7	Marginal
Partial power LF	1.7	14.5	21.8	Band dependent
Partial power MF	18.2	0.2	12.1	Band dependent
Partial power HF	16.6	6.9	1.1	Band dependent
Hit-rate (per minute)	46.0	57.3	80.7	**Unreliable**
Hit-rate (per rotation)	76.0	69.5	64.6	**Unreliable**

Table C.12: Directional repeatability of BTF features (median symmetric relative difference, %).