

Vibration Signatures of Fatigue

Indicator Comparison and Prognostic Assessment
in Resonance Fatigue Testing

Afonso Costa

Vibration Signatures of Fatigue

Indicator Comparison and Prognostic
Assessment in Resonance Fatigue Testing

by

Afonso Costa

Thesis Supervisors: Dr. ir. John-Alan Pascoe, Dr. Vahid Yaghoubi
Project Duration: November, 2025 - September, 2026
Faculty: Faculty of Aerospace Engineering, Delft

Acknowledgements

This thesis began as a study of a quantum photonic vibrometer and, somewhere along the way, became something quite different. I am grateful to the people who made that redirection possible, and even more grateful to those who supported me through all the unexpected turns that came with it.

To my supervisors, Dr. ir. John-Alan Pascoe and Dr. Vahid Yaghoubi, thank you for your guidance, support, and, perhaps most importantly, for giving me the freedom to explore ideas that took this project in directions I had not originally anticipated. I am particularly grateful for the many discussions, questions, and pieces of advice along the way. Your willingness to think through problems with me, challenge my assumptions, and let me find my own way forward had a significant influence on both the research and my experience of doing it.

I would also like to thank DEMO, the laboratory technicians, and the members of the Damage Management & Certification and Structural Dynamics groups. Much of the experimental work in this thesis would simply not have been possible without their help. Thank you for making equipment, materials, facilities, and, just as importantly, your time and expertise available whenever they were needed.

To my friends and fellow students, thank you for the conversations, the distractions, the encouragement, and all the small moments that made the long days of research a little lighter. Thank you for listening to me talk about experiments that did not work, and for reminding me that there was more to life than the thesis.

Finally, to my family, thank you for your constant support, patience, and encouragement throughout this journey. You have been there through the good days and the considerably less good ones, and your support never depended on whether an experiment worked or whether I had any idea what I was doing next. I am deeply grateful for that.

Afonso Costa Delft, September 2026

Abstract

Resonance vibration testing accumulates fatigue cycles far faster than conventional servo-hydraulic methods by driving a specimen at its own natural frequency, and the same response that applies the load also carries information about the damage accumulating within it. However, the literature offers no systematic comparison of competing vibration-based damage indicators measured simultaneously on the same metallic specimens, and no characterisation of how far ahead of failure such indicators support a useful remaining-life estimate.

This thesis addresses both gaps. Eight notched Al 6061-T6 beam specimens were driven to failure at a fixed excitation frequency and interrogated between load blocks with a scanning laser Doppler vibrometer, yielding 33 features per block. A five-layer pipeline ranked those features by sensitivity, combined them into six competing damage indices, and estimated remaining useful life by trend extrapolation and by a population Gaussian Process. Two further specimens were interrupted at the point of detection for optical fractography.

Measurement reliability, rather than indicator choice, governed what the campaign resolved. Only two modes could be tracked stably, mode-shape correlation between two undamaged blocks of the same specimen fell as low as 0.785, and mode labels did not consistently correspond to the same physical mode. Modal parameters changed abruptly at the transition to full-amplitude loading rather than progressively, which physical inspection attributed to the clamping and load-transfer arrangements. Indicators requiring no modal identification, principally the harmonic distortion index and amplitude RMS, proved more dependable than frequency, damping, or coherence, and no damage index reached a resolved population-level prognostic horizon.

These outcomes form a traceable account of where vibration-based fatigue prognosis succeeds, where it does not, and which failures are attributable to method rather than principle. It delivers a ranked indicator comparison, the first prognostic horizon characterisation for this application on metals, quantified design requirements for resonance fatigue fixturing, and three reusable open software tools.

Contents

Acknowledgements	i
Abstract	ii
Nomenclature	xv
1 Introduction	1
1.1 Background	1
1.2 Problem Statement	2
1.3 Context and Scope of This Work	2
1.4 Research Objectives and Questions	3
1.5 Scientific Relevance and Practical Significance	3
1.6 Thesis Outline	3
2 Literature Review	5
2.1 Fatigue Mechanics and Damage Accumulation	5
2.1.1 Phenomenology of Metal Fatigue	5
2.1.2 Fracture Mechanics and Crack Growth	6
2.1.3 Continuum Damage Mechanics Framework	6
2.2 Fundamentals of Structural Dynamics and Vibration	7
2.2.1 Single- and Multi-Degree-of-Freedom Systems	7
2.2.2 The Frequency Response Function	8
2.3 Resonance Vibration Fatigue Testing	9
2.3.1 Operating Principle and Specimen Design	10
2.3.2 Excitation Control: Fixed-Frequency vs. Resonance-Tracked Protocols	10
2.3.3 Block-Cycle Protocol and Interleaved Monitoring	11
2.4 Vibration Signatures of Fatigue Damage	11
2.4.1 Linear Signatures: Natural Frequency, Damping, Amplitude, and Phase	12
2.4.2 Nonlinear Signatures: Crack Breathing and Higher Harmonics	13
2.4.3 Relative Onset and Practical Implications	13
2.5 Vibration Parameters as Damage Indicators	14
2.5.1 Global Indicators: Scalar FRF-Derived Features	15
2.5.2 Spatial Indicators: Mode-Shape-Derived Features	16
2.6 Modal Analysis Techniques	17
2.6.1 Modal Parameter Identification	18
2.6.2 Mode Shape Assembly from Multi-Point Measurements	18
2.6.3 Validation of the Identified Modal Model	19
2.7 Vibration Measurement for Fatigue Monitoring	19
2.7.1 Context: Vibration Sensing Technologies	20
2.7.2 Physical Principles of LDV	20
2.7.3 Noise Sources and Sensitivity Limits	21
2.7.4 Scanning LDV for Mode Shape Acquisition	22
2.7.5 Optical Fractography as an Independent Ground Truth	22
2.8 Data-Driven Damage Detection and Characterisation	23
2.8.1 Baseline Modelling and Multivariate Anomaly Detection	23
2.8.2 Dimensionality Reduction and Indicator Sensitivity Ranking	24
2.8.3 Damage Index Construction	24
2.8.4 Learned Health Indicators: Autoencoder-Based Damage Indices	25
2.9 Remaining Useful Life and Prognosis	26
2.10 Research Gap	27

2.10.1	Gap 1: No Systematic Multi-Indicator Comparison Under Fixed-Frequency Resonance Testing on Metals	27
2.10.2	Gap 2: Absence of a Prognostic Horizon Analysis for Modal-Parameter-Based RUL Estimation on Metals	28
2.10.3	Gap 3: Validating Resonance-Based Testing Against Standard Fatigue Characterisation for Metals	28
2.10.4	Thesis Positioning and Research Questions	29
3	Experimental Design and Development	30
3.1	Synthetic Data Toolkit	31
3.1.1	Physical Model	32
3.1.2	Three-Tier Modal Analysis Hierarchy	32
3.1.3	Sensor Models and Damage Indicators	34
3.1.4	Role in the Experimental Campaign	35
3.2	Specimen Design and Material	37
3.2.1	Material	37
3.2.2	Geometry and Notch Design	38
3.2.3	Specimen Population	39
3.3	Experimental Setup and Instrumentation	43
3.3.1	Fixture and Loading Configuration	44
3.3.2	Scanning Laser Doppler Vibrometer	45
3.3.3	Signal Conditioning and Data Acquisition	46
3.4	Pre-Campaign Characterisation	46
3.4.1	Sensor Characterisation (Tests T1-T4)	47
3.4.2	Specimen and Sample Characterisation (Tests S1-S5)	48
3.4.3	Load Configuration Study	50
3.5	Fatigue Test Protocol	52
3.5.1	Block Structure	52
3.5.2	Phase 3 - Healthy Baseline	53
3.5.3	Phase 4 - Resonance Fatigue Loading	54
3.5.4	Data Storage	54
3.5.5	Independent Validation via Optical Fractography	55
4	Analysis Pipeline Methodology	57
4.1	Signal Processing and Modal Feature Extraction (Layer 0)	58
4.1.1	Indicator Selection Rationale	58
4.1.2	Signal Pre-processing	59
4.1.3	FRF Estimation and Modal Identification	59
4.1.4	Mode Tracking Across Blocks	60
4.1.5	Manual Pole-Selection Review and Modal Identification Quality Checks	60
4.1.6	Feature Vector	61
4.1.7	Anomaly Detection Baseline (Layer 0 Output)	63
4.2	Learned Health Indicators (Layer 0.5)	64
4.2.1	From Raw Scans to Log-FRF Sub-Windows	64
4.2.2	Network Architecture and Resonance-Weighted Loss	64
4.2.3	Two Health Indicators: HI1 and HI2	65
4.3	From Many Indicators to One Damage Index	65
4.3.1	Damage Sensitivity and Redundancy Analysis	65
4.3.2	Principal Component Analysis	66
4.3.3	SHAP-Based Indicator Ranking	67
4.3.4	Multiple Damage Index Formulations	68
4.3.5	Logistic Trend Fitting and Physical Interpretation	69
4.3.6	Inter-Specimen Comparison	70
4.3.7	Comparison with Synthetic Data Toolkit Predictions	70
4.4	Remaining Useful Life Estimation (Layer 3)	70
4.4.1	Failure Threshold	71
4.4.2	Method 1: Parametric Trend Extrapolation	71

4.4.3	Method 2: Gaussian Process Regression	72
4.4.4	Prognostic Horizon and Evaluation Framework	73
4.5	The S-N Curve Module	74
4.6	Live Monitoring	75
4.6.1	Polling and Block Processing	75
4.6.2	Two Alarm Classes	75
5	Results	77
5.1	Specimen Overview and Data Completeness	78
5.2	Modal Parameter Evolution (R1)	79
5.2.1	FRF Evolution	79
5.2.2	Frequency Shift and Anomaly Detection: Automatic Pole Selection	81
5.2.3	Frequency Shift and Anomaly Detection: Manual Pole Selection	82
5.2.4	Feature Evolution	83
5.2.5	Spatial Features: Representative Specimens	86
5.2.6	Modal QA: Cross-Mode MAC Matrices	87
5.3	Indicator Sensitivity Ranking (R2)	89
5.3.1	Damage Sensitivity and Redundancy	89
5.3.2	SHAP-Based Indicator Ranking	90
5.3.3	Ranked Indicator Table	91
5.3.4	Comparison with Layer 0.5 Health Indicators	93
5.4	Damage Index Trajectories (Layer 2)	96
5.4.1	DI Trajectories and Logistic S-Curve Fits	97
5.4.2	Comparison with Synthetic Data Toolkit Prediction	98
5.5	Remaining Useful Life Estimation (R3)	99
5.5.1	RUL Trajectories: Representative Specimens	100
5.5.2	LOO-CV Prediction Accuracy	102
5.6	S-N Characterisation and Resonance Testing Context (R4)	103
5.6.1	FEMAP Stress Calibration	103
5.6.2	Per-Specimen Fatigue Life and Force Characterisation	104
5.6.3	Basquin S-N Fit	104
5.6.4	Prognostic Lead-Time Overlay	106
5.7	Independent Validation via Optical Fractography	106
5.7.1	Live-Monitoring Alarm Summary	107
5.7.2	Stage-by-Stage Optical Evidence	109
5.7.3	Cross-Specimen Summary	113
5.8	Diagnostic Specimen Groups D1 and D2	113
5.8.1	Group D1: Boundary Condition and Load Transfer	113
5.8.2	Group D2: FRF Consistency Under the Revised Configuration	114
6	Discussion	116
6.1	Modal Parameter Evolution Under Resonance Fatigue (R1)	116
6.1.1	Data Quality, Modal Identification Challenges, and Sources of Uncertainty	116
6.1.2	The Expected Sequence: CDM Framework Predictions	119
6.1.3	Frequency, Damping, Coherence, and Peak Amplitude: The Cost of Modal Identification Dependence	120
6.1.4	Phase Indicators: A Mixed Picture	122
6.1.5	Spatial Indicators: Localisation and Mode Shape Distortion	122
6.1.6	Comparison with Finite Element Predictions	123
6.1.7	Diagnostic Evidence from Specimen Groups D1 and D2	124
6.1.8	What R1 Can Still Be Said to Show	124
6.2	Indicator Sensitivity and Ranking (R2)	125
6.2.1	Physical Interpretation of the Ranking	125
6.2.2	Redundancy Structure and Feature Space Geometry	126
6.2.3	Hand-Engineered vs. Autoencoder-Based Indicators	126
6.3	Damage Index Behaviour and CDM Model Consistency	128
6.3.1	Which Formulation Best Represents Damage Progression?	128

6.3.2	Comparison with the Synthetic Data Toolkit Prediction	128
6.4	Remaining Useful Life Estimation (R3)	129
6.4.1	Prognostic Horizon: What It Means in Practice	129
6.4.2	Method 1 vs. Method 2: Trend Extrapolation vs. Gaussian Process	130
6.4.3	Population Size, Data Quality, and the Practical Reliability of RUL Estimates	130
6.5	Resonance Testing as a Monitoring Platform (R4)	130
6.5.1	S-N Characterisation: Method Limitations Before Literature Comparison	130
6.5.2	The Fixed-Frequency Protocol: Amplification vs. Stress Control	132
6.6	Fractography Validation: What the Optical Evidence Says	132
6.6.1	Did the Detection Coincide with Real Damage?	132
6.6.2	Fracture Surface Consistency and Protocol Validity	133
6.6.3	Cross-Specimen Fractographic Evidence for the Fatigue Mechanism	133
6.7	Importance of Findings and Contribution to the Field	134
6.7.1	Closing the Three Literature Gaps	134
6.7.2	Scientific Contribution and Broader Significance	135
7	Conclusions and Recommendations	137
7.1	Summary of Main Findings	137
7.2	Conclusions per Research Question	138
7.2.1	R1 – Vibration Response Evolution	138
7.2.2	R2 – Indicator Sensitivity Ranking	138
7.2.3	R3 – Remaining Useful Life Estimation	138
7.2.4	R4 – Resonance Testing in Context	139
7.2.5	Other Contributions	139
7.3	Practical Recommendations	140
7.3.1	Indicator and Feature Set Selection	140
7.3.2	Damage Index Formulation and Prognosis Method	140
7.3.3	Protocol Design: Fixed-Frequency vs. Tracked-Frequency	141
7.3.4	Live Monitoring and Validation Practice	141
7.3.5	Designing a Vibration Fatigue Test Setup	141
7.4	Limitations of the Study	142
7.4.1	Experimental Campaign Scale	142
7.4.2	Measurement Reliability and Calibration	143
7.4.3	Methodological Scope	143
7.4.4	External Constraints	143
7.5	Suggestions for Future Research	144
7.5.1	Improving Measurement Reliability and Calibration	144
7.5.2	Expanding the Experimental Campaign	144
7.5.3	Methodological Scope: Protocol Comparison, CAE Integration, and Nonlinear Modal Analysis	144
7.5.4	External Constraints: Fractography Depth and the QPV Comparison	145
	References	146
A	Sensor Characterisation Reports (T1–T4)	150
A.1	T1 – Noise Characterisation	151
A.2	T2 – Minimum Detectable Amplitude	152
A.3	T3 – Broadband and Harmonic Excitation Quality	152
A.4	T4 – Disturbance Sensitivity	153
B	Modal Identification Method Selection	155
C	Supplementary Results Figures	156
C.1	Remaining Useful Life Panels: Remaining Formulations	156
C.2	Supplementary Optical Fractography	158
D	Software Toolkit and Code Availability	159
D.1	The Three Tools	159
D.2	Traceability: Thesis Section to Software Module	159

D.3 Documentation and Reproduction	160
Declaration on the Use of Generative AI	161

List of Figures

2.1	Magnitude (top, log scale) and phase (bottom) of the SDOF FRF near resonance, for three damping ratios. Lower damping produces a taller, narrower resonance peak and a steeper phase transition through -90° at $r = 1$ – the property that makes phase deviation a sensitive damage indicator under a fixed-frequency drive (Section 2.4).	9
2.2	Hypothesised, CDM-motivated ordering of damage-indicator onset across the fatigue life, staged around the initiation, propagation, and near-failure phases. This diagram is qualitative: it reflects the expected ordering argued in the text, not a measured timeline, and stage boundaries and onset windows are illustrative rather than calibrated to specific data.	14
2.3	Heterodyne LDV configuration. The laser beam is split into a reference arm, shifted by a fixed carrier frequency f_c via an acousto-optic modulator, and a measurement arm directed at the moving specimen surface. The Doppler-shifted return beam recombines with the reference beam at the photodetector; demodulating the resulting photocurrent around f_c recovers the surface velocity $v(t)$.	21
3.1	Structure of the experimental campaign described in this chapter, with section references. Groups D1 and D2 were added after the main campaign had concluded and are described in Section 3.2.3; their results are reported separately in Section 5.8.	31
3.2	Normalised natural frequency evolution $f_n/f_{n,0}$ for the first four modes over the simulated life, comparing Tier 2 (analytical, dashed) against Tier 3 (FEM, solid). As with Figure 3.5, this comparison predates the final specimen definition and uses the design-phase candidate geometry.	33
3.3	Probability-of-detection analysis for the Polytec PSV-500 (LDV) sensor model, from 200 Monte Carlo trials per point with 95% Clopper-Pearson confidence bounds, plotted against equivalent crack depth (top left), the damage scalar D (top right), fatigue cycle count (bottom left), and broken down by individual indicator (bottom right).	34
3.4	Multi-sensor probability-of-detection comparison (200 Monte Carlo trials per point) for the three modelled sensors, against physical damage size (left), damage scalar D (centre), and cycle count (right).	35
3.5	Tier 3 (FEM) spatial fields at an early damage increment, computed for the design-phase candidate geometry rather than the specimen finally adopted, since the notch profile was not settled until the Group B study (Section 3.4.3): mode 1 shape (top), von Mises stress field (middle), and CDM damage field D (bottom). All three fields are concentrated at the notch, close to the clamped end (left of each panel), rather than at the beam midspan, consistent with the notch placement described in Section 3.2.2.	36
3.6	Tier 3 OAT sensitivity analysis for final damage D_{p95} . Left: tornado chart, showing excitation frequency as the only parameter with non-negligible normalised sensitivity. Right: the underlying effect curve for the top three parameters, showing the sharp resonance window responsible for this result – damage saturates to its maximum value only within a narrow band of excitation frequency, collapsing to zero outside it.	37
3.7	Final specimen geometry (Al 6061-T6, clamped over the first 45 mm): (a) dimensioned drawing; (b) a manufactured specimen. The drawing dimensions the 200 mm beam originally specified; the stock plate ultimately obtained yielded a 196 mm beam, and all reported results refer to that dimension.	39
3.8	Specimens used in the pre-test campaign to explore specimen designs, loading, and the broader experimental setup.	40

3.9	Group D specimens after testing. At $130 \times 20 \times 1$ mm, the beam is substantially shorter than the Group C geometry of Figure 3.7; the notch geometry is unchanged, and the through-hole near the upper end carries the bolted load-transfer connection selected in the Group D1 study. Every specimen shown has separated at the notch.	41
3.10	Schematic and photograph of the experimental setup. The specimen is clamped horizontally at one end; the microshaker applies transverse excitation near the notch, toward the free end, via an instrumented load cell. The scanning LDV measures the out-of-plane velocity response along the beam neutral axis from a fixed stand-off distance.	43
3.11	Revised boundary condition and load transfer arrangement arrived at in the Group D1 study (Section 3.2.3) and used for Group D2, shown for comparison against the main-campaign configuration above. In (b) the specimen is blurred by its own motion during excitation.	44
3.12	S1 baseline modal characterisation: LSCF stabilisation diagram (bottom) with the corresponding FRF magnitude (top). Poles that remain stable in frequency and damping across increasing model order (green) are identified as physical; unstable or new poles at a given order (red crosses) are discarded. Five modes were identified and manually confirmed from this diagram for the specimen shown.	48
3.13	S2 modal parameter uncertainty: tracked natural frequency f_n for each of the five identified modes across repeated short-window estimates, with the mean (solid line) and $\pm\sigma_f$ band (shaded) shown per mode. Frequency uncertainty σ_f varies by mode, from sub-Hz for the lowest, well-separated mode to several Hz for closely-spaced higher modes, illustrating why the anomaly detection threshold must be calibrated per mode rather than applied as a single global value.	49
3.14	S4 nonlinearity validation: LDV velocity response PSD at the drive frequency, showing distinct super-harmonic peaks at $2f$, $3f$, $4f$, and beyond, well above the noise floor. The presence and relative amplitude of these super-harmonics is what the HDI (Equation 2.13) quantifies; their persistence across several harmonics at this drive level is the basis for the reported nonlinearity onset.	50
3.15	Load configuration study: EMA mobility (top) and accelerance (middle) magnitude, and force coherence γ_F^2 (bottom), for repeated runs at a candidate configuration. Apart from the first mode, the 0.85 coherence threshold (dashed) is met across the first several bending modes for both the LDV and reference accelerometer channels, confirming adequate FRF estimation quality at this configuration.	51
3.16	Per-specimen test protocol. Phase 3 (healthy baseline, 5 blocks at sub-endurance-limit amplitude) and Phase 4 (fatigue loading at target amplitude, run to fracture; block count varies by specimen) share an identical block structure. The normalised cycle fraction $\tau = N/N_f$ is computed post-hoc once N_f is known from fracture.	52
3.17	Structure of a single fatigue block. The harmonic loading step accumulates damage and provides instantaneous frequency, amplitude, and HDI features; the broadband scanning step extracts the full modal feature vector (point tier and spatial tier) that enters the five-layer analysis pipeline.	53
4.1	Structure of the analysis methodology described in this chapter. Raw block data feeds two independent indicator-construction tracks (Sections 4.1 and 4.2), which converge on a shared sensitivity-ranking and damage-index step (Section 4.3) feeding remaining-useful-life estimation (Section 4.4). S-N characterisation (Section 4.5) and live monitoring/validation (Section 4.6) are independent of this chain, requiring only the raw force/cycle records and the real-time data stream respectively.	58
5.1	FRF magnitude evolution for each of the eight Group C1 specimens (S01–S08, corresponding to C1-01–C1-08), averaged across scan points, coloured from the first Phase 3 block (blue) to the final recorded Phase 4 block (red). Dotted vertical lines mark the resonance search-band seeds established during pole identification.	80
5.2	Automatic pole selection: natural frequency f_n (Mode 0), Hotelling T^2 statistic, Harmonic Distortion Index, and CUSUM accumulator, all eight Group C1 specimens, harmonic-loading-stage data. Shaded region: Phase 3 baseline blocks.	81

5.3	Representative N4SID stabilisation diagram used for manual pole-selection review (specimen S05, Phase 4, block 7), showing stable poles (green), frequency-stable-only poles (orange), and unstable/noise poles (red) across model orders, overlaid on the H1 FRF magnitude. The automatic selection is marked with a dashed vertical line for comparison.	82
5.4	Manual pole selection: natural frequency f_n (Mode 0), Hotelling T^2 statistic, Harmonic Distortion Index, and CUSUM accumulator, all eight Group C1 specimens, harmonic-loading-stage data, poles assigned via manual review rather than automatic tracking.	83
5.5	Feature evolution, part 1 of 2 (Features 1–12 of 22), all eight Group C1 specimens vs. block number. See Table 4.2 for feature definitions.	84
5.6	Feature evolution, part 2 of 2 (Features 13–22 of 22), all eight Group C1 specimens vs. block number.	85
5.7	Spatial amplitude (left) and curvature (right) deviation fields, Mode 0: specimen S04 (top row) and specimen S05 (bottom row); scan point index vs. block number, coloured by deviation in σ from the first Phase 3 block. Dashed vertical line: Phase 3/Phase 4 boundary.	86
5.8	Spatial amplitude (left) and curvature (right) deviation fields, Mode 1: specimen S04 (top row) and specimen S05 (bottom row); scan point index vs. block number, coloured by deviation in σ from the first Phase 3 block.	87
5.9	MAC matrices for a sample of specimen and block combinations, exported from the pipeline's modal QA tool. Top-left: block-versus-block comparison of two healthy Phase 3 baseline blocks of specimen S01. Bottom-right: cross-mode comparison for a Phase 4 fatigue block of specimen S06. Values beyond the 0.95 mode-mixing threshold are flagged in red by the source tool.	88
5.10	Median Spearman rank correlation $ \rho_s $ between each of the 33 Layer 0 features and the cycle fraction τ , aggregated across the eight Group C1 specimens. Orange: point-tier (Tier 1) features; yellow: spatial-tier (Tier 2) features. Hatched bars: features flagged as redundant ($ r_{ij} > 0.90$ with a higher-ranked partner).	89
5.11	Global SHAP importance Φ_j^{global} for each of the 33 Layer 0 features, sorted in descending order. Orange: point-tier features; yellow: spatial-tier features.	90
5.12	Early-life SHAP importance Φ_j^{early} ($\tau < 0.2$, relaxed to $\tau < 0.4$ for specimens with fewer than 20 blocks) for each of the 33 Layer 0 features, sorted in descending order. Orange: point-tier features; yellow: spatial-tier features.	91
5.13	Layer 0.5 health indicators (64-dimensional bottleneck): (left) HI1, resonance-weighted reconstruction MSE; (right) HI2, maximum latent z -score deviation. Both panels show all eight Group C1 specimens vs. scan block, with the Phase 3 99th-percentile threshold shown as a dashed horizontal reference.	93
5.14	CAE validation via leave-one-specimen-out cross-validation (64-dimensional bottleneck). Left: ROC curve of HI1 and HI2 against the independent T^2 /Isolation Forest/CUSUM detector, with chance-level performance (AUROC = 0.50) shown as a dashed diagonal. Right: per-specimen false-positive rate (HI1, own Phase 3 blocks) for the CAE trained on the other seven specimens only, against a nominal 1% target derived from the 99th-percentile threshold definition.	94
5.15	Layer 0.5 health indicators for the 16-dimensional bottleneck ($\alpha = 5$) and 32-dimensional bottleneck ($\alpha = 15$) configurations, for comparison against Figure 5.13.	95
5.16	LOSO-CV validation for the 32-dimensional bottleneck configuration, for comparison against Figure 5.14.	96
5.17	Damage index trajectories $D(\tau)$ for all six formulations, all eight Group C1 specimens (S01–S08, consistent colour coding with Figures 5.1–5.2). Solid lines: raw per-block $D(\tau)$. Dashed lines: fitted logistic S-curve per specimen. Grey shaded band: Phase 3 baseline reference range. Dotted horizontal line: failure threshold $D_f = 0.90$.	97
5.18	Synthetic Data Toolkit's predicted damage evolution for a representative excitation case (1.2 mm amplitude, 6 cm drive-point location). Top left: damage index D_{95} vs. cycles. Top right: natural frequency ratio $f_n/f_n(0)$ for all six simulated modes. Bottom left: effective stiffness ratio E_{eff}/E_0 . Bottom right: modal damping ratio ζ for all six modes.	99

5.19	RUL trajectory (Layer 3 native export), D_{freq} formulation, all eight Group C1 specimens overlaid. Top left: RUL trajectory (true RUL, dotted; Method 1 fitted curve with 95% CI; Method 2 GP checkpoint estimates at $\tau = 0.2, 0.4, 0.6$ with error bars), tick marks show each specimen's τ_h . Top right: CI width vs. τ against the $\alpha = 30\%$ prognostic-horizon threshold. Bottom left: raw $D_{\text{freq}}(\tau)$ with failure threshold $D_f = 0.90$	100
5.20	RUL trajectory (Layer 3 native export), D_{T^2} formulation, all eight Group C1 specimens overlaid. Panel layout identical to Figure 5.19.	101
5.21	RUL trajectory (Layer 3 native export), D_{PC1} formulation, all eight Group C1 specimens overlaid. Panel layout identical to Figure 5.19.	101
5.22	FEMAP static finite-element analysis for specimen C1-03, elemental Von Mises stress contour with exaggerated predicted displacement deformation overlaid. The notch root, where the peak stress used to derive $k_{\text{specimen}} = \sigma_{\text{FEM}}/F_{\text{FEM}}$ is extracted, is visible at the transition between the clamped section (left) and the reduced-section gauge length.	104
5.23	Basquin S-N characterisation of the eight Group C1 specimens (log-log axes), with the Layer 3 prognostic lead-time overlay. Filled circles: per-specimen failure data (S_a , N_f). Solid line: OLS Basquin fit $N = C \cdot S_a^{-b}$. Shaded region: 95% prediction band (individual observations). Triangles: prognostic lead-time markers for the DI formulations that reached τ_h for at least one specimen; horizontal arrows point from the lead-time marker to the specimen's actual failure point, with the lead time in absolute cycles labelled. S_a values use the per-specimen k calibration (Table 5.6).	105
5.24	Specimen C2-01 (dashboard label S09, Table 5.7) – live-monitoring record across Phase 4 blocks up to interruption. Top to bottom: T^2 statistic; HDI (dashed line: 3σ threshold); CAE HI1 and CAE HI2 (dashed line: Phase 3 99th-percentile threshold, displayed but excluded from the alarm trigger); Amplitude RMS. Right: final-block indicator summary (reproduced in Table 5.8).	107
5.25	Specimen C2-02 (dashboard label S10, Table 5.7) – live-monitoring record across Phase 4 blocks up to interruption. Panel layout identical to Figure 5.24.	108
5.26	Specimen C2-01 – profilometer export of the notch-root region at Stage 1, pre-test (top row), and Stage 3, at interruption (bottom row). Both rows were acquired at the same magnification over an identical 18.09×24.128 mm field of view, and show, left to right: the optical image, the height map, the 3D reconstruction, and that row's height colour scale. The scales are auto-scaled independently per row, so colours are not comparable between them.	109
5.27	Specimen C2-01 – close-up optical image of the notch-root region at the point of interruption, showing the surface condition referenced in answering Q1.	110
5.28	Specimen C2-02 – close-up optical image of the notch-root region at the point of interruption.	110
5.29	Specimen C2-01 – optical scans of the notch-root region at three stages: (left) Stage 1, pre-test; (centre) Stage 3, at the point of interruption; (right) Stage 4, the fracture surface after loading was resumed to failure. The right-hand panel views the fully separated fracture face edge-on, with the stable-growth region and the demarcated final-fracture zone both visible; the material surrounding it is intact beam, not an unbroken ligament. Panels are rendered with photographic surface texture rather than height colour-mapping, and are not to a common scale.	112
5.30	Fracture-surface comparison: Group C1 reference specimens (labelled “Sample 1” and “Sample 5” in the source export) shown alongside the corresponding 3D height-mapped scans, used as the Group C1 reference for the Stage 4 comparison against specimens C2-01 (Figure C.5 (Appendix C.2)) and C2-02 (Figure C.6 (Appendix C.2)).	112
5.31	H_1 FRF progression for Group D2 specimens D2-02 and D2-03, complex-averaged over nine scan points, all runs overlaid. Top: magnitude. Centre: phase. Bottom: ordinary coherence γ^2 , with the shaded band showing the spread across scan points.	114
5.32	H_1 FRF progression for Group D2 specimens D2-04 and D2-05. Panel layout identical to Figure 5.31.	115

6.1	Clamping arrangement used for the Group C1 campaign, at the fixed end of the specimen. The jaw faces of the machine vice carry a stepped and grooved profile intended for gripping round or irregular stock, and scoring and edge damage from prior use are visible on the jaw surfaces.	117
6.2	Fractured accelerometer mounting arm, Specimen 7. The fracture pattern is not consistent with a component subjected only to pure back-and-forth transverse motion, and is offered here as direct visual evidence for a mixed bending-and-axial loading history at the shaker attachment point (discussed in the text, and further in Section 6.5.1).	118
6.3	Per-specimen HDI evolution through Phase 4 for three representative specimens (S02, S04, S08), each shown on its own scale. All three hold at a low, near-constant plateau for most of their recorded life, with a rise confined to the final block or two before fracture; the magnitude of that final rise differs by more than two orders of magnitude between S02/S04 and S08.	121
A.1	T1 noise characterisation, shaker off, for both the LDV and the reference accelerometer on the load cell. (a) Noise PSD: the LDV noise floor sits several orders of magnitude below the accelerometer's across the full bandwidth. (b) Cumulative RMS noise from 0 to f : the LDV reaches approximately 1.0×10^{-5} m/s within the 0–1000 Hz band spanning the first three bending modes, rising to approximately 1.8×10^{-5} m/s across the full 50 kHz measurement bandwidth, consistently below the accelerometer's cumulative noise at every frequency.	151
A.2	T2 cross-frequency metric comparison (accelerometer vs. LDV): SNR, amplitude uncertainty, frequency uncertainty, stiffness-change resolution, and THD at 0.1, 10, and 1000 Hz, at the lowest achievable shaker drive voltage. The extreme values at 0.1 Hz (SNR near –300 dB, THD near 200%) reflect that the shaker was driven at a fixed <i>voltage</i> rather than a fixed force or displacement (Section 3.4.3): at 0.1 Hz the resulting displacement is far too small to produce a usable response at all, whereas by 10–1000 Hz the same drive voltage yields an adequately excited, well-behaved response for both sensors.	152
A.3	T3a broadband excitation quality, repeated over three acquisitions under white-noise excitation at 0.1 amplitude. Each column shows, top to bottom: FRF magnitude with the H_1/H_2 spread band and the H_v geometric mean, ordinary coherence γ^2 against the 0.9 “good” and 0.7 “warning” thresholds, and the corresponding noise-to-signal ratio. Coherence is close to unity below approximately 5 kHz and falls away above roughly 10 kHz, which bounds the usable band for FRF estimation well above the 0–500 Hz range within which the campaign's tracked modes lie.	152
A.4	T3b total harmonic distortion against instantaneous frequency along a swept-sine excitation, at two drive amplitudes, for the reference accelerometer (left) and the LDV (right). The 5% distortion threshold is marked. Both sensors remain below the threshold across the low-frequency region containing the campaign's drive frequency, with distortion rising sharply only above approximately 100 Hz for the accelerometer and above 1 kHz for the LDV.	153
A.5	T4a sine-excitation power spectral density measured by the LDV at three optical power settings, plotted against linear (left) and logarithmic (right) frequency axes, with the 10 Hz excitation marked. The measured response at the excitation frequency is unchanged across settings; optical power affects the broadband noise floor rather than the signal, and the effect is confined to frequencies well above the campaign's measurement band.	153
A.6	T4d noise characterisation by specimen surface material, comparing aluminium against CFRP for both the accelerometer and the LDV. Clockwise from top left: noise PSD on linear and logarithmic frequency axes, band-limited RMS by frequency band, and overlapping Allan deviation against averaging time. The LDV's noise floor on the aluminium surface is roughly two orders of magnitude below the accelerometer's across every band, which is the basis for the decision not to apply a surface treatment to the specimens (Section 3.4.1).	154
C.1	RUL trajectory (Layer 3 native export), D_{comp} formulation, all eight Group C1 specimens overlaid. Panel layout identical to Figure 5.19.	156

C.2	RUL trajectory (Layer 3 native export), D_{CAE} (reconstruction) formulation, all eight Group C1 specimens overlaid. Panel layout identical to Figure 5.19.	157
C.3	RUL trajectory (Layer 3 native export), $D_{\text{CAE,latent}}$ formulation, all eight Group C1 specimens overlaid. Panel layout identical to Figure 5.19.	157
C.4	Specimen C2-01 – Stage 1 pre-test surface-roughness characterisation of the notch root (profilometer export). Reported values: S_a (arithmetic mean height) and S_z (maximum height) across sampled areas at the notch root, together with the 3D height-mapped surface. This is the undamaged reference against which the Stage 3 comparison (Figure 5.26) is read.	158
C.5	Specimen C2-01 – Stage 4 fracture surface, formal profilometer export (optical, height-mapped, and 3D reconstruction views).	158
C.6	Specimen C2-02 – Stage 4 fracture surface, formal profilometer export (optical, height-mapped, and 3D reconstruction views).	158

List of Tables

3.1	Material properties of Al 6061-T6 used in the Synthetic Data Toolkit and for experimental data interpretation (MatWeb, LLC, 2026).	38
3.2	Summary of specimen groups, their purpose, and their role in the experimental campaign.	42
3.3	Summary of instrumentation used in the experimental campaign.	46
3.4	Key outputs of the pre-campaign characterisation (S1–S5) and their role in the fatigue test protocol and analysis pipeline.	50
3.5	Key datasets in the per-specimen HDF5 file and their roles in the analysis pipeline.	55
4.1	Mapping of chapter sections onto the research questions they address.	57
4.2	Layer 0 feature families, their extraction domain, and the physical mechanism they are intended to capture.	63
5.1	Mapping of chapter sections onto the research questions they address.	77
5.2	Overview of the ten Group C specimens: Phase-4 (fatigue) cycles to fracture and data availability across this chapter. N_f excludes the five-block Phase 3 baseline common to every specimen and reflects the pipeline’s own precise within-block fracture detection (Section 5.6) rather than a block-count estimate. Layer 0.5 was run for all eight Group C1 specimens.	78
5.3	Ranked indicator table: all 33 Layer 0 features, sorted by composite rank. Spearman ρ is the median across the eight Group C1 specimens and is reported <i>signed</i> , so a feature that decreases monotonically with damage carries a negative value; the ranking itself uses $ \rho $. SHAP columns are blank where the feature was excluded from the gradient-boosted model as redundant with a higher-ranked partner. Composite rank is the mean of the three rank positions (Spearman $ \rho $, global SHAP, early-life SHAP), so tied and skipped rank values are expected.	92
5.4	Logistic S-curve fit parameters across the eight Group C1 specimens for all six DI formulations, transcribed from the pipeline’s own export. Values are median (IQR) over converged fits only. τ_{mid} : inflection point, the cycle fraction at 50% of failure-stage damage. β : steepness parameter, the damage acceleration rate.	97
5.5	Leave-one-specimen-out cross-validated RMSPE for both remaining-life estimation methods, at three evaluation checkpoints τ_c . Method 1 produces no estimate at $\tau_c = 0.2$ for any formulation, since too few blocks are available at that point to support a parametric fit.	103
5.6	Per-specimen force characterisation and fatigue life for the eight Group C1 specimens. F_a : force amplitude extracted from the first Phase 4 block. R : stress ratio. k : per-specimen FEM calibration factor (MPa/N, Section 5.6.1). $S_a = kF_a$. N_f : total Phase-4 cycles to fracture (Table 5.2).	104
5.7	Group C2 specimen protocol summary. $N_{\text{interrupt}}$: absolute cycle count at interruption (Phase 4 only). $N_{f,\text{total}}$: final fracture cycle count after resumption (Phase 4 only), from Table 5.2.	108
5.8	Live-monitoring dashboard, final block before interruption: indicator values against their Phase 3-derived thresholds, for both Group C2 specimens. Bold: indicator exceeded its threshold at this block. CAE HI1/HI2 are displayed by the dashboard but excluded from the alarm-triggering logic (see text).	108
5.9	Cross-specimen summary of the Group C2 optical fractography validation. All entries are qualitative observations from the microscope image record and the live-monitoring dashboard.	113
D.1	Mapping between analysis stages, the thesis sections that specify and report them, and the corresponding pipeline components.	160

Nomenclature

Abbreviations

Abbreviation	Definition
AUROC	Area Under the Receiver Operating Characteristic curve
ASTM	American Society for Testing and Materials
CAE	Convolutional Autoencoder
CDI	Curvature Damage Index
CDM	Continuum Damage Mechanics
CI	Confidence Interval
CUSUM	Cumulative Sum (control chart)
DAF	Dynamic Amplification Factor
DI	Damage Index
DOF	Degree of Freedom
EMA	Experimental Modal Analysis
FE	Finite Element
FEM	Finite Element Method
FRF	Frequency Response Function
GP	Gaussian Process
HCF	High-Cycle Fatigue
HDF5	Hierarchical Data Format version 5
HDI	Harmonic Distortion Index
HI1	Health Indicator 1 (autoencoder reconstruction error)
HI2	Health Indicator 2 (autoencoder latent-space deviation)
IQR	Interquartile Range
ISO	International Organization for Standardization
LCF	Low-Cycle Fatigue
LDV	Laser Doppler Vibrometer
LOO-CV	Leave-One-Out Cross-Validation
LOSO-CV	Leave-One-Specimen-Out Cross-Validation
LSCF	Least-Squares Complex Frequency-domain (modal identification)
MAC	Modal Assurance Criterion
MDOF	Multi-Degree-of-Freedom
MSE	Modal Strain Energy
MSE-TVD	Modal Strain Energy Total Variation Distance
N4SID	Numerical Algorithms for Subspace State-Space System Identification
OAT	One-At-a-Time (sensitivity analysis)
OMA	Operational Modal Analysis
PC1	First Principal Component
PCA	Principal Component Analysis
PHM	Prognostics and Health Management
POD	Probability of Detection
PSD	Power Spectral Density
PSV-500	Polytec PSV-500 scanning laser Doppler vibrometer
QPV	Quantum Photonic Vibrometer
RFP	Rational Fraction Polynomial (modal identification)
RMS	Root Mean Square
RMSPE	Root Mean Square Percentage Error
ROC	Receiver Operating Characteristic

Abbreviation	Definition
RUL	Remaining Useful Life
SDOF	Single-Degree-of-Freedom
SHAP	SHapley Additive exPlanations
SHM	Structural Health Monitoring
SLDV	Scanning Laser Doppler Vibrometer
SSI-COV	Covariance-driven Stochastic Subspace Identification
TVD	Total Variation Distance

Symbols

Symbol	Definition	Unit
a	Crack length or equivalent crack depth	mm
b	Basquin exponent. Two conventions appear: $N_f = (\sigma_a/\sigma'_f)^{1/b}$ in Section 2.1.1, and $N = C S_a^{-b}$ in Chapters 4 and 5	—
C	Basquin coefficient; Paris law coefficient	—
$C(\tau)$	Parameter covariance matrix of the nonlinear least-squares fit	—
CDI_i	Curvature damage index for mode i	—
D	Continuum damage variable; damage index	—
D_f	Failure threshold on the damage index ($D_f = 0.90$)	—
D_{T^2}	Damage index from the Hotelling T^2 statistic	—
D_{freq}	Damage index from normalised frequency shift	—
D_{PC1}	Damage index from the first principal component score	—
D_{comp}	Composite, SHAP-weighted damage index	—
D_{CAE}	Damage index from autoencoder reconstruction error	—
$D_{\text{CAE,latent}}$	Damage index from autoencoder latent-space deviation	—
E	Young's modulus	GPa
$\tilde{E}$	Effective (damaged) Young's modulus	GPa
$F(t)$	Applied harmonic force	N
F_a	Force amplitude	N
F_m	Mean force	N
f	Frequency	Hz
f_c	Acousto-optic modulator carrier frequency	Hz
f_n	Natural frequency	Hz
$f_{n,0}$	Baseline (undamaged) natural frequency	Hz
f_{drive}	Fixed excitation frequency during Phase 4	Hz
$H(\omega)$	Frequency response function	(m/s)/N
H_1	H_1 FRF estimator	(m/s)/N
h	CUSUM decision threshold	—
I	Second moment of area	mm ⁴
K	Stress intensity factor	MPa $\sqrt{\text{m}}$
ΔK	Stress intensity factor range	MPa $\sqrt{\text{m}}$
K_t	Elastic stress concentration factor	—
k	Force-to-stress calibration factor; CUSUM reference value	MPa/N
L	Cantilevered specimen length	mm
MAC_i	Modal assurance criterion for mode i	—
m	Paris law exponent	—
N	Cycle count	cycles
N_b	Total number of blocks in the pooled feature matrix	—
N_f	Cycles to failure	cycles
n_{levels}	Number of distinct stress levels detected by the gap rule	—
p	Number of features in the feature vector	—

Symbol	Definition	Unit
r	Frequency ratio ω/ω_n	—
R^2	Coefficient of determination	—
RUL	Remaining useful life	cycles
S_a	Nominal stress amplitude	MPa
T^2	Hotelling T^2 statistic	—
t	Time; specimen thickness	s; mm
$v(t)$	Surface velocity	m/s
$\mathbf{X}$	Pooled feature matrix	—
α	EMA smoothing factor; significance level; horizon threshold	—
β	Logistic S-curve steepness parameter	—
γ^2	Ordinary coherence	—
Δ_{CI}	Width of the 95% confidence interval on the RUL estimate	—
ζ	Modal damping ratio	—
λ	Laser wavelength	nm
ℓ	Gaussian process kernel lengthscale	—
ρ	Material density	kg/m ³
ρ_s	Spearman rank correlation coefficient	—
σ	Stress	MPa
σ_a	Stress amplitude	MPa
σ'_f	Fatigue strength coefficient	MPa
σ_f	Standard deviation of the identified natural frequency	Hz
σ_ζ	Standard deviation of the identified damping ratio	—
σ_ϕ	Spatial standard deviation of phase across scan points	deg
τ	Cycle fraction, N/N_f	—
τ_c	Evaluation checkpoint in cycle fraction	—
τ_f	Predicted cycle fraction at failure	—
τ_h	Prognostic horizon	—
τ_{mid}	Logistic S-curve inflection point	—
Φ_j^{global}	Global SHAP importance of feature j	—
Φ_j^{early}	Early-life SHAP importance of feature j	—
ϕ	Phase angle	deg
$\Delta\phi_{diff}$	Phase deviation from the healthy baseline	deg
ψ_i	Mode shape vector for mode i	—
ω	Angular frequency	rad/s
ω_n	Undamped natural angular frequency	rad/s

Specimen and Test Nomenclature

Label	Definition
Group A	Preliminary specimens used for geometry and notch exploration
Group B	Specimens used for the iterative load configuration study
Group C1	Main fatigue campaign, eight specimens run to fracture
Group C2	Interrupted-protocol pair used for optical fractography validation
Group D1	Diagnostic specimens for the boundary condition and load transfer study
Group D2	Diagnostic specimens re-run under the revised configuration
S01–S08	Analysis-pipeline labels corresponding to specimens C1-01 to C1-08
T1–T4	Pre-campaign sensor characterisation tests (Appendix A)
S1–S5	Pre-campaign sample characterisation tests
Phase 3	Healthy baseline blocks at sub-endurance excitation
Phase 4	Fixed-frequency resonance fatigue loading to fracture
R1–R4	Research questions (Section 1.4)

Introduction

1.1. Background

Fatigue is the dominant failure mode in metallic structures subjected to cyclic mechanical loading. It accounts for the majority of in-service structural failures across aerospace, mechanical, and civil engineering. Its characteristic insidiousness lies in the fact that damage accumulates progressively and largely invisibly, at the microstructural level, far below the load magnitudes that would cause immediate yielding, until a critical crack length is reached and fracture occurs (Schijve, 2008). The engineering response to this challenge has historically been preventive: conservative design margins, scheduled inspection intervals, and component retirement well before the statistical end of useful life. While effective from a safety standpoint, this approach is wasteful and, increasingly, insufficient as structures and components grow more complex and operational demands become more varied. The ability to detect and track damage in real time, rather than merely to design around its eventual occurrence, is becoming both technically and economically compelling.

Vibration-based structural health monitoring (SHM) offers a route to this capability. The dynamic response of a structure is directly governed by its stiffness and mass distribution, so as fatigue damage accumulates and local stiffness degrades around a growing crack, that response changes in characteristic and, in principle, measurable ways.

Two classes of indicators can be extracted from such a response, and the distinction between them matters throughout this thesis. *Modal* indicators, the natural frequencies, damping ratios, and mode shapes, require a modal identification step: the measured response must first be decomposed into individual modes before any indicator can be computed. *Response-level* indicators are taken from the measured signal directly, without that decomposition, and include the harmonic content generated by a breathing crack, the response amplitude, and the driving force. The two classes carry different information and, as this thesis shows, differ substantially in how reliably that information survives to the measurement. The appeal of the approach is that the measurement infrastructure required, a vibration excitation source and one or more dynamic response sensors, is relatively simple, can be operated continuously and non-intrusively, and produces data that is directly connected to the structural mechanics of the degradation process (Liang et al., 2024; Slavič, 2021).

Resonance vibration fatigue testing is, independently of its role as a monitoring platform, also an attractive alternative to conventional constant-amplitude fatigue testing on servo-hydraulic or electromechanical rigs, which are intrinsically slow: reaching the 10^6 – 10^7 cycles required for a single high-cycle-fatigue data point can take days to weeks on such rigs, a limitation that vibration-based resonance methods have been developed to address for two decades and continue to be refined today (Furman et al., 2025; George et al., 2004). By driving a specimen at or near one of its natural frequencies, resonance testing exploits the structure's own dynamic amplification to reach large cyclic stress amplitudes with modest actuator forces, so that the same number of cycles can be accumulated in hours or even minutes. Česnik et al. (Česnik et al., 2012) demonstrated one million load cycles completed in 21 minutes on an aluminium specimen under phase-controlled resonance excitation. In the laboratory context, res-

onance vibration fatigue testing combines this speed advantage with the SHM ideas introduced above into a single integrated framework. The resonant excitation simultaneously creates the fatigue damage and provides the dynamic signal from which modal parameter evolution can be continuously monitored, so that the same test that accelerates fatigue-life determination also supplies the measurement signal on which damage detection and prognosis depend.

The scanning Laser Doppler Vibrometer (LDV), a non-contact, high-resolution optical instrument, enables both single-point scalar tracking and spatially resolved mode shape acquisition across the same measurement campaign, without any mechanical interference with the specimen. This combination of resonance fatigue testing, scanning LDV measurement, and data-driven modal parameter analysis creates a monitoring framework that is, at once, experimentally mature in its individual components and underexplored in its integrated application to the problem of remaining fatigue life estimation (Tomasini & Castellini, 2020).

1.2. Problem Statement

Despite the physical plausibility of vibration-based fatigue monitoring and the availability of suitable experimental tools, three specific gaps limit its practical applicability:

- an indicator characterisation gap, since the many candidate damage indicators proposed in the literature have rarely been compared systematically under the resonance protocol and material class considered here;
- a prognosis gap, since converting a detected damage trend into a reliable remaining-life estimate requires a continuous damage index, an extrapolation model, and a validated prognostic horizon that has not yet been demonstrated for vibration-based monitoring of metallic specimens;
- a protocol-characterisation gap, since the fixed-frequency resonance protocol's own behaviour as a monitoring platform, relative to tracked-frequency alternatives and conventional fatigue testing, has not been quantitatively assessed.

Chapter 2 develops all three gaps in full, situates them against the relevant published work, and shows how they motivate the four research questions stated in Section 1.4 below.

1.3. Context and Scope of This Work

This thesis was initiated as part of a project to benchmark a quantum photonic vibrometer (QPV) for fatigue testing applications, with the aim of comparing its sensitivity to damage signatures against a conventional LDV under controlled resonance fatigue conditions. The experimental infrastructure, electrodynamic shaker, scanning LDV, data acquisition system, and a structured testing protocol were designed and partially developed in that context.

The QPV instrument became unavailable during the project due to an equipment failure that could not be resolved within the available timeframe. Rather than abandoning the experimental work, the project was reoriented: the fatigue monitoring problem itself, which the original project had treated as a backdrop for a sensor comparison, became the primary scientific focus. The existing infrastructure, the synthetic data toolkit developed for theoretical predictions, and the signal processing and modal analysis pipeline were all retained and expanded to address the monitoring problem directly.

The research questions that emerge from the reorientation are scientifically well-motivated in their own right, address genuine gaps in the literature (detailed in Chapter 2), and are well-matched to the available experimental capabilities. The original QPV comparison remains a natural and clearly signposted direction for future work, should the instrument become available.

The study is scoped to metallic specimens, specifically thin aluminium beam specimens, under fixed-frequency harmonic resonance excitation in a laboratory setting. Composite specimens were initially included in the experimental plan but proved difficult to fracture within the available actuator capacity. Results are therefore reported for metals only. The analysis employs scanning LDV as the sole response measurement instrument; the role of contact sensors is limited to the load cell integrated with the shaker fixture.

The main campaign comprises eight specimens run to fracture, supported by a two-specimen track

tested under a deliberately interrupted protocol to provide optical fractography at the point of detection, and by two further diagnostic groups tested after the main campaign to re-examine the fixture arrangement. Section 3.2.3 describes the full population. A population of this size supports the within-specimen comparisons on which R1 and R2 depend, but constrains the population-level statistical claims available for R3 and R4; this constraint is stated where it bears on each result and consolidated in Section 7.4.

1.4. Research Objectives and Questions

The overarching objective of this thesis is to determine whether, and to what degree, the evolution of a specimen's vibration response during resonance fatigue testing of metallic specimens can be used to detect damage, characterise its progression, and estimate remaining fatigue life. Both classes of indicator introduced in Section 1.1 are considered: modal indicators requiring an identification step, and response-level indicators taken directly from the measured signal. This objective is addressed through four research questions, whose full motivation against the literature is developed in Chapter 2 (Section 2.10):

- R1** *How does the vibration response evolve as fatigue damage accumulates under resonant vibration loading?*
- R2** *Which vibration-based indicators are most sensitive to early fatigue damage?*
- R3** *Can vibration response trends be used to estimate remaining fatigue life?*
- R4** *How does resonance-based vibration fatigue testing compare to traditional fatigue tests?*

1.5. Scientific Relevance and Practical Significance

The scientific contribution of this thesis is an integrated, empirically grounded characterisation of vibration-based fatigue prognosis under controlled resonance conditions, spanning both modal and response-level indicators, a contribution that is, to the author's knowledge, not available in the existing literature for metallic specimens. The ranked indicator table produced in response to R2 and the prognostic horizon analysis produced in response to R3 are intended as directly usable guidance for the design of future fatigue monitoring systems: which indicators to prioritise, and from what point in a specimen's life a prognostic estimate can be trusted. The independent optical-fractography validation track (Section 3.5.5) provides a further, vibration-independent check on the detection logic underlying this prognosis. As Chapter 5 reports, several of these questions are answered in the negative under the conditions achieved in this campaign, and a substantial part of the contribution lies in establishing where the limits fall and which of them are attributable to the method rather than to the principle.

The practical significance extends to structural health monitoring in aerospace and mechanical engineering more broadly. Resonance vibration fatigue testing is widely used in qualification and endurance-trial campaigns for aerospace components (Di Maio & Magi, 2015; George et al., 2004). Demonstrating that the test itself, with minimal additional instrumentation, can yield continuous remaining life estimates, and do so within a fraction of the time required by conventional fatigue methods (Section 1.1), would represent a meaningful step toward condition-based maintenance in laboratory settings, and a proof of concept for in-service monitoring applications.

1.6. Thesis Outline

The remainder of this thesis is structured as follows.

Chapter 2, Literature Review surveys the theoretical and experimental foundations relevant to this study, covering fatigue mechanics, resonance fatigue testing methodology, vibration signatures of damage, both modal and response-level parameters as damage indicators, modal analysis techniques, LDV measurement, and data-driven prognosis methods. The chapter concludes with a formal statement of the research gap that motivates the four research questions above.

Chapter 3, Experimental Design and Development describes the synthetic prediction toolkit, specimen design, instrumentation, and pre-campaign characterisation, culminating in a comparison between predicted and measured modal behaviour that justifies the final specimen population, force levels, and

test protocol actually used in the campaign. It also introduces the diagnostic specimen groups D1 and D2, added after the main campaign to re-examine the boundary-condition and load-transfer arrangements.

Chapter 4, Analysis Pipeline Methodology describes the analysis built on that experimental foundation: two independently constructed indicator sets, hand-engineered vibration indicators spanning both classes and a learned, autoencoder-based health indicator, which converge on a shared sensitivity-ranking and damage-indexing step feeding the remaining-useful-life estimation used to address R1–R3; alongside two further, independent investigations, an S-N characterisation module that answers R4 directly, and a live-monitoring and optical-fractography validation track that provides an independent, physically observed check on the detection logic.

Chapter 5, Results presents the experimental outcomes for all research questions, including the vibration response evolution records (R1), the indicator sensitivity ranking for both the hand-engineered and autoencoder-based indicators (R2), the damage index trajectories and prognostic horizon analysis (R3), the S-N characterisation and resonance testing context (R4), the independent live-monitoring and optical-fractography validation of the detection logic, and the diagnostic evidence obtained from specimen groups D1 and D2.

Chapter 6, Discussion interprets the results in the context of the literature and the pipeline parameter settings underlying them, reviews the validation evidence, and examines the study's contribution and its implications for future monitoring system design and for the broader resonance fatigue testing methodology.

Chapter 7, Conclusions and Recommendations summarises the main findings, states the conclusions with respect to each research question, identifies the study's contribution and limitations, and outlines directions for future work, including the QPV comparison that was the original motivation for this project.

Three software tools were developed for this work: the LDV Analysis Suite used for pre-campaign characterisation, the Synthetic Data Toolkit used for physics-informed prediction, and the five-layer analysis pipeline, described in Chapters 3 and 4 respectively. All three are publicly available and separately documented; Appendix D describes each, and maps every analysis stage reported in this thesis to the component that implements it.

2

Literature Review

2.1. Fatigue Mechanics and Damage Accumulation

Fatigue is a progressive and localised damage mechanism that develops in materials and structures subjected to cyclic loading. It is one of the primary causes of structural failure in engineering systems, often occurring at stress levels substantially below the material's static yield strength. Since fatigue damage initiates at the microstructural level and evolves gradually over long operational periods, its early detection poses significant challenges for conventional inspection techniques (Schijve, 2008). Improving the ability to detect and track damage at an early stage is therefore essential for extending service life, reducing maintenance costs, and ensuring structural safety.

This section establishes the physical foundations required to interpret and model the fatigue damage evolution observed in the experimental campaign. The focus is placed on the mechanics of metallic fatigue, covering the phenomenological S-N characterisation, the fracture mechanics description of crack growth, and the continuum damage mechanics (CDM) framework used to translate these mechanisms into trackable, quantitative damage states.

2.1.1. Phenomenology of Metal Fatigue

Fatigue life in metallic materials is classically divided into three stages: crack initiation, stable crack propagation, and unstable final fracture (Lee, 2005; Schijve, 2008). During crack initiation, cyclic plastic slip at stress concentrations, such as surface irregularities, notches, or microstructural heterogeneities, leads to the formation of persistent slip bands and, eventually, micro-cracks. This stage typically dominates the total fatigue life in high-cycle fatigue (HCF) regimes, where applied stresses are largely elastic and failure requires millions of cycles. In contrast, low-cycle fatigue (LCF) regimes involve significant plastic deformation per cycle, and crack propagation consumes a proportionally larger fraction of the total life.

Depending on the applied strain amplitude and the number of cycles to failure N_f , fatigue tests are commonly classified into LCF ($N_f \lesssim 10^4$) and HCF ($N_f \gtrsim 10^5$) regimes. Resonant vibration fatigue testing, the experimental approach employed in this thesis, operates primarily in the HCF regime, where the elastic nature of the response and the well-defined modal behaviour of the specimen make continuous dynamic monitoring practically feasible. The implications of this regime choice are discussed further in Section 2.3.

The phenomenological relationship between applied stress amplitude σ_a and fatigue life N_f is described by the Basquin (power-law) S-N relationship:

$$N_f = \left(\frac{\sigma_a}{\sigma'_f} \right)^{1/b} \quad (2.1)$$

where σ'_f is the fatigue strength coefficient and b is the Basquin exponent (typically -0.05 to -0.15 for metallic alloys) (Lee, 2005; Schijve, 2008). The S-N curve is defined in terms of fully reversed loading ($R = -1$). For non-zero mean stresses, corrections such as the Goodman or Morrow relations are applied. A material-specific feature of the S-N curve for ferrous metals is the endurance limit σ_{end} , a stress amplitude below which fatigue failure is not expected within a conventionally defined life (commonly 10^6 cycles for aluminium alloys). The endurance limit serves as a practical design threshold and is the quantity used to anchor the Basquin model within the synthetic data toolkit developed alongside this thesis.

The Basquin formulation of Equation 2.1 is widely accepted as an adequate description of macroscopic S-N behaviour for metallic materials in the HCF regime (Lee, 2005; Schijve, 2008). Disagreement in the literature centres on how well such simple power-law models capture life scatter, load-sequence effects, and mean-stress dependency under variable-amplitude loading. These complexities are secondary to the present experimental context, which employs single-amplitude harmonic excitation, so the Basquin model is adopted as the baseline for life estimation.

2.1.2. Fracture Mechanics and Crack Growth

Once a fatigue crack has nucleated, its propagation can be described within the framework of linear elastic fracture mechanics (LEFM). The central parameter is the stress intensity factor range ΔK , which characterises the amplitude of the crack-tip stress field:

$$\Delta K = Y \Delta \sigma \sqrt{\pi a} \quad (2.2)$$

where $\Delta \sigma$ is the applied stress range, a is the current crack half-length, and Y is a geometry-dependent correction factor. The empirical Paris law relates crack growth rate to the ΔK of Equation 2.2:

$$\frac{da}{dN} = C (\Delta K)^m \quad (2.3)$$

where C and m are material constants (Schijve, 2008). The Paris law of Equation 2.3 captures the stable crack propagation regime, where growth is approximately linear in log-log space. It underpins the characteristic S-shaped damage index curve expected in experiments: crack growth is slow during initiation, accelerates as ΔK increases with crack length (since $\Delta K \propto \sqrt{a}$), and becomes unstable as K_{max} approaches the fracture toughness K_{Ic} .

This S-shaped progression has direct implications for the structure of the data analysis pipeline. The damage indices constructed in this analysis framework (Section 2.8) are expected to follow a logistic trend consistent with Paris-law dynamics, and the prognostic horizon (the cycle fraction at which reliable remaining-life estimates first become available) is closely related to the inflection point of this curve.

2.1.3. Continuum Damage Mechanics Framework

While fracture mechanics provides a physically rigorous description of crack growth, its application requires explicit crack geometry information that is rarely available during continuous monitoring. Continuum damage mechanics (CDM) offers a complementary formulation in which damage is represented by a scalar field variable $D \in [0, 1]$, where $D = 0$ denotes the pristine undamaged state and $D = 1$ corresponds to complete material failure (Lemaitre & Chaboche, 1990; Murakami, 2012; Silitonga et al., 2013). This scalar representation is adopted as the primary damage-tracking framework in this thesis precisely because it requires no explicit crack-geometry measurement, aligning with the goal of estimating damage state from vibration response alone; fracture mechanics is retained as the physical basis for interpreting why the CDM-predicted S-shaped trajectory arises, not as a candidate for direct online tracking. In these approaches, the effective Young's modulus degrades as:

$$E(D) = E_0 (1 - \alpha D) \quad (2.4)$$

where E_0 is the undamaged modulus and α is a material-specific stiffness degradation coefficient (typically $\alpha \approx 1.5$ for aluminium alloys). The degradation law of Equation 2.4 directly links the CDM damage

variable to the measurable quantity of interest in this thesis: since natural frequencies scale approximately as $\omega_n \propto \sqrt{k/m} \propto \sqrt{E}$, progressive stiffness degradation translates into monotonically decreasing natural frequencies. A first-order approximation gives:

$$\frac{\Delta\omega_n}{\omega_{n,0}} \approx -\frac{1}{2} \frac{\Delta k}{k_0} \approx -\frac{\alpha}{2} D \quad (2.5)$$

This relationship motivates the use of relative frequency shift as a primary damage index and provides a physical basis for expecting a monotonic, approximately S-shaped evolution of $D(t)$ over the specimen's fatigue life. However, it should be noted that other effects may change this relationship. For example, Lorenzino and Navarro (Lorenzino & Navarro, 2015) showed that for aluminium specimens, the natural frequency can initially rise slightly due to material hardening before monotonically decreasing as crack area grows.

The CDM framework also accommodates spatial distributions of damage. In the experimental setup, damage is expected to localise at the highest-stress region of the specimen. As damage localises, its effect on global modal parameters diminishes with increasing mode order unless the damaged region coincides with a mode shape anti-node. This spatial selectivity is exploited in the mode shape curvature and the Modal Assurance Criterion (MAC) indicators discussed in Section 2.5.

The key point of agreement in the literature is that, regardless of the chosen theoretical framework, fatigue damage produces a gradual, progressive reduction in structural stiffness that is, in principle, detectable through changes in dynamic response (Liang et al., 2024; Montalvão et al., 2006; Slavič, 2021). The practical challenge, and one of the central questions motivating this thesis, is not whether such changes exist, but whether they can be detected early enough, and characterised accurately enough, to enable reliable remaining fatigue life estimation.

2.2. Fundamentals of Structural Dynamics and Vibration

The preceding section established why fatigue damage should, in principle, be visible in a structure's dynamic response. This section makes that link concrete by setting out, in one place, the structural dynamics vocabulary and equations that the rest of the chapter draws on repeatedly: the equations of motion for single- and multi-degree-of-freedom systems, the mode shape vector and its role as the basis for spatial damage indicators, and the Frequency Response Function together with its magnitude, phase, and practical estimation from measured data. The Dynamic Amplification Factor, the specific consequence of this machinery that makes resonance fatigue testing practical, is introduced where it is used, in Section 2.3.

2.2.1. Single- and Multi-Degree-of-Freedom Systems

The simplest model of a vibrating structure is the single-degree-of-freedom (SDOF) oscillator: a mass m , spring k , and viscous damper c , whose motion $x(t)$ under an applied force $f(t)$ obeys

$$m\ddot{x}(t) + c\dot{x}(t) + kx(t) = f(t). \quad (2.6)$$

Two derived quantities characterise the free (unforced) response and recur throughout this thesis: the undamped natural frequency $\omega_n = \sqrt{k/m}$, and the damping ratio $\zeta = c/(2\sqrt{mk})$, which expresses the damping level relative to the critical damping at which oscillation just ceases (Rao, 2017). For the lightly damped metallic specimens considered here ($\zeta \approx 0.001$ – 0.005), the system oscillates many times before the free response decays appreciably, which is precisely the behaviour that resonance fatigue testing exploits.

A real structure is more accurately represented as a multi-degree-of-freedom (MDOF) system, discretised into N coordinates and described by mass, damping, and stiffness matrices $\mathbf{M}$, $\mathbf{C}$, $\mathbf{K} \in \mathbb{R}^{N \times N}$:

$$\mathbf{M}\ddot{\mathbf{x}}(t) + \mathbf{C}\dot{\mathbf{x}}(t) + \mathbf{K}\mathbf{x}(t) = \mathbf{f}(t). \quad (2.7)$$

Solving the associated undamped eigenvalue problem $(\mathbf{K} - \omega_r^2 \mathbf{M})\phi_r = \mathbf{0}$ yields N natural frequencies ω_r , and their corresponding mode shapes ϕ_r , the spatial displacement patterns associated with each resonance (Ewins, 2000; Rao, 2017). For proportionally damped systems, the mode shapes diagonalise Equation 2.7 into a set of N decoupled SDOF equations, one per mode, each governed by Equation 2.6 with its own modal mass, modal damping ratio ζ_r , and modal stiffness. This modal decomposition is what justifies treating each resonance of a real specimen, in isolation, using SDOF theory. Throughout this thesis, quantities such as natural frequency shift and damping ratio increase are tracked mode-by-mode using exactly this SDOF machinery, even though the specimen itself is a continuous, many-mode structure.

Because the mode shape vector ϕ_r is only defined up to an arbitrary scale factor, a raw element-by-element comparison between a measured shape and its undamaged baseline is not meaningful on its own; a consistent scaling convention (e.g. mass-normalisation) fixes this ambiguity for computing derived modal quantities, but comparing two shapes measured at different times still requires a scale-invariant measure. This motivates the use of normalised correlation metrics, most notably the Modal Assurance Criterion (MAC) introduced in Section 2.5, which quantify shape similarity independently of any common scaling. The mode shape vector itself is not read directly from a single sensor: it must be reconstructed from response measurements taken at multiple points, a procedure detailed in Section 2.6.2 (Allemang & Avitabile, 2022; Ewins, 2000).

2.2.2. The Frequency Response Function

The Frequency Response Function (FRF) $H(\omega)$ is the frequency-domain solution of Equation 2.6: the ratio of the steady-state response to a unit-amplitude harmonic force at frequency ω . Depending on which response quantity is used, the FRF is termed receptance (displacement/force), mobility (velocity/force), or accelerance (acceleration/force). The three are related by factors of $i\omega$ and carry the same information. For the SDOF system of Equation 2.6, the receptance FRF is

$$H(\omega) = \frac{1}{k - m\omega^2 + ic\omega} = \frac{1}{k} \cdot \frac{1}{(1 - r^2) + i(2\zeta r)}, \quad r = \frac{\omega}{\omega_n}, \quad (2.8)$$

a complex-valued function whose magnitude and phase are examined next. For an MDOF system, the FRF between excitation point j and response point k is a sum of SDOF-like terms, one per mode, each weighted by the product of the mode shape components at j and k . Near an isolated resonance, the FRF is dominated by the single nearby mode and Equation 2.8 remains an excellent local approximation. This is the basis for the SDOF local identification (circle-fit) method described in Section 2.6.1.

The phase of Equation 2.8, the angle by which the response lags the applied force, is

$$\phi(r) = -\arctan\left(\frac{2\zeta r}{1 - r^2}\right). \quad (2.9)$$

Figure 2.1 plots the magnitude of Equation 2.8 and the phase of Equation 2.9 against r for three damping ratios. Well below resonance ($r \ll 1$) the response is in phase with the force ($\phi \approx 0^\circ$), while at resonance ($r = 1$) the response lags by exactly 90° regardless of damping; well above resonance ($r \gg 1$) the response is in anti-phase ($\phi \rightarrow -180^\circ$). The phase slope $d\phi/dr$ is steepest exactly at $r = 1$, which is what makes phase such a sensitive detuning indicator whenever a structure is driven near one of its natural frequencies while that frequency is itself drifting, a point taken up again in Section 2.4 in the context of fatigue-induced damage detection.

In practice, $H(\omega)$ is never observed directly: it must be estimated from finite, noise-contaminated time records using spectral averaging. The most widely used estimator is the H_1 estimator:

$$\hat{H}_1(\omega) = \frac{S_{FV}(\omega)}{S_{FF}(\omega)} \quad (2.10)$$

where S_{FV} is the cross-power spectral density between force and velocity, and S_{FF} is the auto-power spectral density of the force, both computed via Welch's method using overlapping Hann-windowed

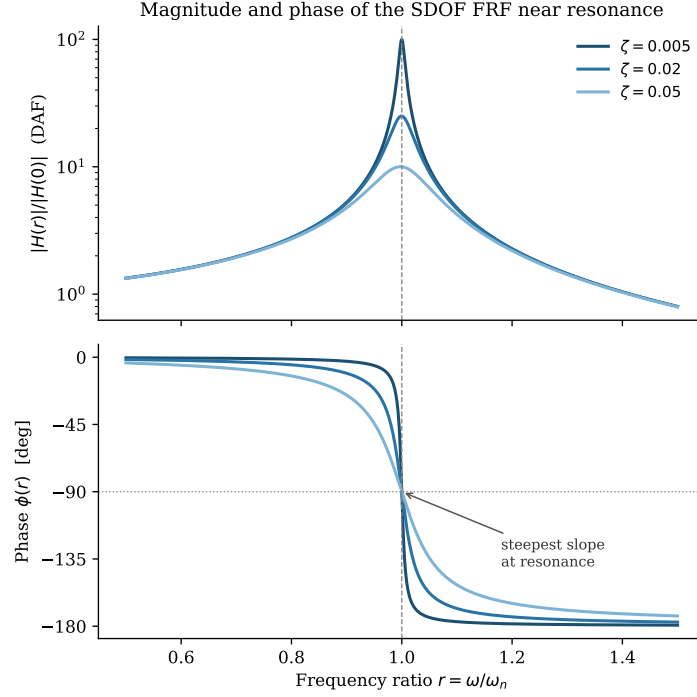


Figure 2.1: Magnitude (top, log scale) and phase (bottom) of the SDOF FRF near resonance, for three damping ratios. Lower damping produces a taller, narrower resonance peak and a steeper phase transition through -90° at $r = 1$ – the property that makes phase deviation a sensitive damage indicator under a fixed-frequency drive (Section 2.4).

segments (Allemang & Avitabile, 2022; Bendat & Piersol, 1993; Ewins, 2000; Harris & Piersol, 2002; Welch, 1967). The H_1 estimator of Equation 2.10 minimises the effect of noise on the output channel. The alternative H_2 estimator minimises input noise, and is preferred when the force measurement is noisier than the response. In shaker-based modal testing using an LDV, the response measurement is often very clean, but H_1 remains the standard estimator in most practical applications because the force measurement typically has a sufficiently high signal-to-noise ratio, and H_1 provides robust FRF estimates under common experimental conditions.

The coherence function $\gamma^2(\omega)$ accompanies the FRF estimate as a quality metric:

$$\gamma^2(\omega) = \frac{|S_{FV}(\omega)|^2}{S_{FF}(\omega) \cdot S_{VV}(\omega)} \quad (2.11)$$

Coherence near unity at resonance indicates that the velocity is linearly explained by the measured force. Drops in coherence signal nonlinearity, poor signal-to-noise ratio, or, in fatigue monitoring, the onset of crack-breathing nonlinearity (Allemang & Avitabile, 2022; Bendat & Piersol, 1993; Ewins, 2000). The coherence of Equation 2.11 thus serves simultaneously as a data quality gate and as an independent early indicator of structural change.

The spectral resolution of the FRF, and therefore the precision of subsequent frequency estimates, is governed by the segment length used in the Welch average. For a sampling rate of f_s and a segment of N_{seg} samples, the frequency resolution is $\Delta f = f_s/N_{\text{seg}}$. For lightly damped specimens with narrow resonances and small expected frequency shifts (of order 0.1–1%), the segment length must be chosen to place Δf well below the expected shift magnitude, typically requiring segments of several seconds duration (Allemang & Avitabile, 2022).

2.3. Resonance Vibration Fatigue Testing

Conventional fatigue testing relies on servo-hydraulic or electromechanical actuators to apply cyclic loads at frequencies far below the specimen's natural frequencies, following standardised procedures

such as ASTM E466 and ISO 1099 (ASTM International, 2021; International Organization for Standardization, 2017). While this approach offers direct control over the applied force or displacement, it is intrinsically slow: for high-cycle fatigue (HCF) regimes requiring 10^6 to 10^7 cycles, test durations of weeks or months are common. Resonance vibration fatigue testing addresses this limitation by exploiting dynamic amplification, driving a structure near one of its natural frequencies allows large cyclic stress amplitudes to be generated with comparatively small excitation forces (Česnik et al., 2012; George et al., 2004; Slavič, 2021).

This section describes the operating principles of resonance fatigue testing, its practical implementation with an electrodynamic shaker, and the specific test protocol employed in this thesis.

2.3.1. Operating Principle and Specimen Design

In a resonance fatigue test, a specimen is mounted on the armature of an electrodynamic shaker, typically as a cantilever beam or notched plate, and excited by a harmonic base-excitation signal at or near its fundamental natural frequency ω_n .

The magnitude of the FRF (Equation 2.8), normalised by the static deflection $1/k$ the same force would produce, is the Dynamic Amplification Factor (DAF):

$$\text{DAF}(r) = \frac{1}{\sqrt{(1-r^2)^2 + (2\zeta r)^2}}. \quad (2.12)$$

At resonance ($r = 1$), Equation 2.12 reduces to $1/(2\zeta)$. This dynamic amplification is what makes resonance fatigue testing practical: for the lightly damped metallic specimens considered here ($\zeta \approx 0.001$ – 0.005), the DAF reaches values of 100–500, meaning the stress amplitudes in the fatigue zone are two to three orders of magnitude larger than those the shaker force alone would produce statically. This amplification makes it possible to cycle a specimen to failure in minutes rather than days at frequencies of several hundred hertz, while remaining in the HCF regime where linear elastic behaviour and well-defined modal properties are preserved (Česnik et al., 2012; Di Maio et al., 2018; Furman et al., 2025; Teng et al., 2022).

A central requirement of resonance fatigue testing is that the fatigue zone, meaning the region intended to accumulate damage and eventually fracture, must coincide with the location of maximum dynamic stress in the chosen mode shape. For a cantilever beam under first-mode bending excitation, the stress is maximised at the clamped root; for a notched specimen, the notch is positioned at the expected stress concentration. This co-location of damage zone and stress maximum ensures that, as damage accumulates, the consequent changes in local stiffness are reflected in the global modal parameters, making them useful damage indicators throughout the fatigue life (Česnik et al., 2012; Di Maio et al., 2018).

Specimen geometry also governs the test frequency, which, through the DAF, directly determines the rate of cycle accumulation. A higher natural frequency is generally desirable for throughput, but must be balanced against actuator bandwidth, load cell resolution, and the requirement that the modal response be well-separated from higher modes to avoid response complexity. (Česnik et al., 2012; Di Maio et al., 2018; Magi et al., 2016b).

2.3.2. Excitation Control: Fixed-Frequency vs. Resonance-Tracked Protocols

A fundamental consideration in the design of a resonance fatigue test is how the excitation frequency is managed as the specimen's natural frequency drifts during damage accumulation. Two broad strategies exist in the literature.

- In a *resonance-tracked* protocol, the excitation frequency is continuously adjusted, typically via a phase-locked loop, to maintain the drive frequency at or near the instantaneous natural frequency of the specimen. This ensures a constant Dynamic Amplification Factor and, consequently, a constant stress amplitude in the fatigue zone throughout the test. Česnik et al. (Česnik et al., 2012) demonstrated this approach on aluminium specimens, using the phase angle between excitation and response as a real-time control signal. The phase is -90° at resonance, and

deviations from this value trigger automatic corrections to the drive frequency. This methodology enables accurate S-N characterisation because the applied stress remains constant to within a few percent throughout the specimen's life. Modern phase-locked-loop controllers for this kind of resonance tracking continue to be refined for robustness and speed (Hippold et al., 2024).

- In a *fixed-frequency* protocol, the shaker drives the specimen at a constant excitation frequency set equal to the initial natural frequency established in the baseline characterisation phase. As damage accumulates and the natural frequency drifts generally downward (due to progressive stiffness loss, as described in Section 2.1), the specimen operates slightly off-resonance. The response amplitude decreases as the resonance peak shifts away from the fixed drive frequency, which means the applied stress is no longer constant. The fixed-frequency approach trades load control accuracy for measurement simplicity. Since the excitation frequency is invariant, consecutive frequency response measurements are directly comparable without normalisation for excitation frequency variation, and phase-based damage indicators retain a consistent physical reference (Di Maio et al., 2018; Lorenzino & Navarro, 2015).

2.3.3. Block-Cycle Protocol and Interleaved Monitoring

Building on the two excitation strategies of Section 2.3.2, resonance fatigue tests can be organised in two ways. A continuous test, most naturally paired with a resonance-tracked protocol, holds the specimen in resonance throughout and is well suited when the only goal is detecting a frequency shift (Česnik et al., 2012; Magi et al., 2016b).

The alternative, used in this thesis, organises testing as a sequence of alternating loading and measurement phases. During each fatigue block, the specimen is excited at the harmonic drive frequency for a defined number of cycles N_{block} , accumulating damage. At the end of each block, the harmonic excitation is briefly replaced by a broadband frequency sweep, allowing Frequency Response Functions (FRFs) to be measured across the full frequency range of interest. Modal parameters, natural frequencies, damping ratios, and, where a scanning LDV is used, mode shapes, are extracted from these FRFs and compared to the baseline established before fatigue loading commenced (Di Maio et al., 2018; Magi et al., 2016b).

This block-cycle interleaved protocol offers two key advantages. First, it provides a continuous record of modal parameter evolution throughout the specimen's fatigue life, enabling the construction of damage indices as a function of consumed life fraction (Section 2.8). Second, by separating the high-amplitude harmonic loading phase from the low-amplitude broadband characterisation phase, it preserves measurement linearity: the modal parameters are extracted in a regime where the nonlinear effects associated with crack breathing are suppressed, giving cleaner and more interpretable FRF data. The trade-off is that the damage state is sampled discretely at the block boundaries rather than continuously, introducing a time-resolution limitation.

2.4. Vibration Signatures of Fatigue Damage

The physical link between fatigue damage and dynamic response was established in Section 2.1: progressive stiffness degradation, whether described through CDM or fracture mechanics, shifts the natural frequencies of the structure and alters the spatial distribution of modal strain energy. This section examines how these physical changes manifest in measurable vibration signals, with a focus on the specific signatures observable in metallic specimens under harmonic resonance excitation.

Vibration signatures of fatigue damage can be broadly categorised as *linear* or *nonlinear*. The first appear in the conventional Frequency Response Function (FRF) and are fully consistent with a linear, time-invariant model of the structure, encompassing frequency, damping, amplitude, and phase, all read from the same complex quantity. The second arise from the contact mechanics of crack faces, which violate the linearity assumption. Each category has a distinct physical origin, a characteristic time of onset during the fatigue life, and different implications for detection sensitivity and measurement requirements.

2.4.1. Linear Signatures: Natural Frequency, Damping, Amplitude, and Phase

Four distinct linear signatures are read from the same underlying quantity, the complex FRF near resonance (Section 2.2.2): its pole location gives the natural frequency, its imaginary part gives damping, its magnitude gives resonance amplitude, and its angle gives phase. Each arises from a distinct physical mechanism, described in turn below; their practical use as monitoring features, including estimation methods and reported quantitative thresholds, is addressed separately in Section 2.5.

Natural frequency shift.

The most widely studied linear signature of fatigue damage is the downward shift in natural frequency. As a fatigue crack propagates, local material is effectively removed from the load path, reducing the structure's effective stiffness k_{eff} at the damage site. Since natural frequencies scale as $\omega_n \propto \sqrt{k_{\text{eff}}/m}$, a stiffness reduction produces a monotonically decreasing frequency for a fixed modal mass. As established by the CDM relationship in Section 2.1, this decrease follows an approximately S-shaped trajectory: slow during the initiation-dominated early life, accelerating through the propagation phase, and asymptoting as failure approaches and the net section collapses (Česnik et al., 2012; Liang et al., 2024; Slavič, 2021).

An important nuance, documented experimentally by Lorenzino and Navarro (Lorenzino & Navarro, 2015), is that natural frequency can exhibit a slight increase in the very early fatigue life before the monotonic decrease begins. This initial rise is attributed to cyclic hardening: plastic deformation at the crack tip increases the local dislocation density and temporarily stiffens the material before crack area growth becomes the dominant mechanism. For aluminium alloys in particular, this hardening-then-softening sequence is experimentally reproducible and constitutes a distinguishable feature in the frequency evolution record. The transition from the hardening to the softening regime correlates closely with the onset of measurable crack propagation and can therefore serve as a marker for the end of the initiation phase.

The spatial sensitivity of frequency shifts depends on the relationship between the damage location and the mode shape. A crack at a displacement anti-node (where strain energy density is maximum) produces the largest frequency perturbation for a given crack size. Conversely, damage near a node, where modal curvature is low, may produce negligible frequency changes even for relatively large cracks. This mode-shape dependence motivates monitoring multiple modes simultaneously: a crack that is poorly positioned relative to one mode may be well-sensed by another, and the pattern of differential frequency shifts across modes can, in principle, be inverted to localise the damage (Liang et al., 2024; Montalvão et al., 2006; Singh & Sehgal, 2022).

Modal damping.

Fatigue damage introduces additional energy-dissipation mechanisms: friction between crack faces under cyclic loading, thermoelastic coupling at the crack tip, and dislocation dragging in the plastic zone, which collectively increase the effective damping ratio ζ_n of the affected modes. Because these mechanisms are active from the earliest stages of microcrack formation, rather than requiring a macroscopic crack to have already formed, damping is expected to respond earlier in the fatigue life than natural frequency shift, a point developed further in Section 2.4.3.

Resonance amplitude.

At resonance, the FRF magnitude is approximately $|H(\omega_n)| \approx 1/(2\zeta_n\omega_n^2 m)$. An increase in damping therefore reduces the peak amplitude even if mass and stiffness are unchanged, while a reduction in stiffness raises it; in practice the two effects act simultaneously, so resonance amplitude is a composite signature of both underlying mechanisms rather than an independent one.

Phase deviation.

The phase of the FRF, introduced in Equation 2.9 (Section 2.2.2), evolves as a function of the frequency ratio $r = f_{\text{drive}}/f_n$. When $f_{\text{drive}} = f_n$ (exact resonance), $\phi = -90^\circ$ regardless of damping. As f_n decreases due to damage while f_{drive} remains fixed (the fixed-frequency protocol of this thesis), r decreases below unity, and the phase moves away from -90° in a predictable manner. Because the phase slope $d\phi/df$ is steepest exactly at resonance (Section 2.3.1), a

small frequency shift produces a disproportionately large phase perturbation whenever the drive remains near-resonant, which is the physical reason phase deviation is expected to be a comparatively sensitive damage signature; the empirical evidence for this, and the practical use of phase deviation as a monitoring feature, are discussed in Section 2.5.

2.4.2. Nonlinear Signatures: Crack Breathing and Higher Harmonics

When a fatigue crack is open throughout the loading cycle, the structure can be treated as linear with a reduced effective stiffness. However, when the cyclic stress field causes the crack faces to alternately open and close, the so-called breathing crack mechanism, the effective stiffness of the structure becomes a function of the instantaneous displacement. This bilinear or piecewise-linear stiffness variation acts as a nonlinear element in the otherwise linear system, generating response components at integer multiples of the excitation frequency (super-harmonics) even under purely sinusoidal excitation (Andreus & Baragatti, 2012; Proso et al., 2016; Worden et al., 2002; Worden et al., 2008).

Harmonic Distortion Index.

The Harmonic Distortion Index (HDI) is a direct measure of this nonlinearity, defined as the ratio of super-harmonic power to the power at the fundamental drive frequency:

$$\text{HDI} = \frac{\sum_{k=2}^K P(k f_{\text{drive}})}{P(f_{\text{drive}})}, \quad (2.13)$$

where $P(f)$ is the power spectral density at frequency f and K is the number of super-harmonics included. In an undamaged specimen the HDI is near zero, rising only once crack-face contact becomes significant. Because HDI responds to the contact mechanics of the crack faces rather than to the stiffness reduction that drives the linear signatures above, it is largely immune to the confounds affecting those indicators: a specimen with a through crack that remains open throughout the cycle shows strong frequency shift but weak harmonic distortion, whereas a specimen with a partially-closed surface crack may show the opposite pattern. Its practical use as a monitoring feature, including the measurement sensitivity this requires, is discussed in Section 2.5.

Amplitude-dependent frequency shift.

A related nonlinear effect is amplitude-dependent frequency shift: the instantaneous resonance frequency of a specimen with a breathing crack varies with the response amplitude, because the proportion of the cycle during which the crack is closed depends on the stress ratio. This means that sweeping the excitation amplitude at fixed frequency produces a characteristic bending of the resonance peak, a signature detectable through bispectral or higher-order spectral analysis (Carminati & Ricci, 2018; Worden et al., 2002).

2.4.3. Relative Onset and Practical Implications

A recurring theme in the literature is the question of which signature appears first in the fatigue life, and therefore which indicator offers the longest prognostic horizon. The Continuum Damage Mechanics framework (Section 2.1) motivates a specific physical ordering for this question, staged around the three phases of fatigue life. In the *initiation stage*, before a macroscopic crack has formed, crack nucleation and early microcrack coalescence increase internal friction and energy dissipation at the grain scale; damping ratio and the Harmonic Distortion Index are the indicators expected to respond first, since damping is driven by crack-face friction and thermoelastic dissipation at the crack tip, while HDI reflects the crack-breathing nonlinearity that activates as soon as a partially-closed crack is present (Andreus & Baragatti, 2012; Mei et al., 2019). Phase deviation at the fixed drive frequency is expected to amplify this same early sensitivity, since it responds directly to the detuning between the drive frequency and the shifting instantaneous resonance: even a sub-threshold frequency shift produces a disproportionately large phase deviation at the steepest point of the phase curve (Di Maio et al., 2018; Lorenzino & Navarro, 2015). In the *propagation stage*, once a dominant macrocrack grows under cyclic loading, stiffness loss accelerates, producing the monotonic natural frequency decline that CDM theory links to damage accumulation through $\Delta\omega_n/\omega_{n,0} \approx -\frac{\alpha}{2}D$ (Equation 2.5). In the *near-failure stage*, mode shape

distortion local to the crack becomes detectable, and the spatial indicators introduced in Section 2.5, the Modal Assurance Criterion, modal curvature, and mode shape energy redistribution, are expected to show their largest changes.

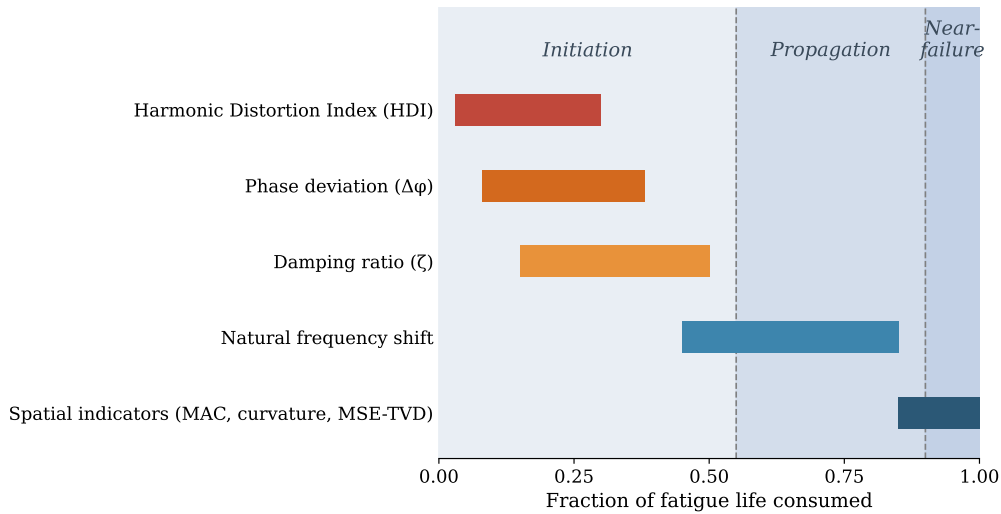


Figure 2.2: Hypothesised, CDM-motivated ordering of damage-indicator onset across the fatigue life, staged around the initiation, propagation, and near-failure phases. This diagram is qualitative: it reflects the expected ordering argued in the text, not a measured timeline, and stage boundaries and onset windows are illustrative rather than calibrated to specific data.

This CDM-motivated staging (Figure 2.2) is broadly consistent with the empirical consensus reported across the experimental literature, though the ordering is not uniformly supported: nonlinear signatures (HDI, amplitude-dependence) → phase deviation → damping increase → natural frequency shift → spatial mode-shape distortion (Andreaus & Baragatti, 2012; Di Maio et al., 2018; Liang et al., 2024; Magi et al., 2016b; Worden et al., 2002).

Natural frequency shift tends to be one of the last global indicators to reach detectable levels, but is also the most robust and reproducible. Damping and phase occupy a useful intermediate position: more sensitive than frequency shift, but more stable than nonlinear indicators. However, a key nuance is that the relative ordering is not universal and depends on the material, specimen geometry, loading amplitude, and measurement system. In particular, whether nonlinear signatures genuinely precede linear ones in metallic specimens under harmonic resonance loading, rather than being masked by noise, remains an open experimental question that the present campaign is positioned to contribute to.

2.5. Vibration Parameters as Damage Indicators

Section 2.4 established the physical mechanisms by which fatigue damage produces measurable changes in vibration response, including the sensitivity/monotonicity tension and typical onset ordering discussed in Section 2.4.3: a useful indicator must be *sensitive* (detectable above the measurement noise floor well before failure) and *monotonic* (trend-consistent over the fatigue life), and the two properties tend to trade off against each other. This section explores what indicators the field has proposed and what is known about each one's sensitivity, monotonicity, and practical limitations, so that a full multi-indicator feature set can be assembled for the analysis pipeline of Section 2.8.

Following the global/spatial framing introduced at the end of Section 2.4, indicators are organised here into two tiers. Global indicators are scalar quantities derived from single-point FRF measurements. They carry information about the aggregate dynamic state of the specimen. Spatial indicators are vector or field quantities derived from multi-point scanning measurements. They carry information about where damage is located and how it distorts the mode shape.

A second distinction cuts across that one and is developed in parallel throughout this section, because it proves consequential later. *Modal* indicators, comprising natural frequency, damping ratio, and every mode-shape-derived quantity, can only be computed once the measured response has been decom-

posed into individual modes, so their reliability is bounded by the reliability of that identification step. *Response-level* indicators, comprising the harmonic distortion index, response amplitude, and driving force, are computed from the measured signal directly and require no such decomposition. An indicator can therefore be global without being modal: the harmonic distortion index below is a scalar single-point quantity, and so global in the tier sense, but it involves no modal identification at all.

2.5.1. Global Indicators: Scalar FRF-Derived Features

Natural frequency shift.

The fractional shift in natural frequency, $\Delta f_n / f_{n,0} = (f_{n,0} - f_n(t)) / f_{n,0}$, is the most widely used and best-understood damage indicator in vibration-based fatigue monitoring (Andreus & Baragatti, 2009; Česnik & Slavič, 2014; Česnik et al., 2012; Liang et al., 2024; Lorenzino & Navarro, 2015). As a monitoring feature its practical advantages are considerable: natural frequency can be estimated from a short FRF measurement with high precision, and the resulting time series is smooth, largely free of measurement outliers, and readily interpretable.

Its primary limitation as an early indicator is its *late onset*. Experimental studies of metallic specimens consistently show that detectable frequency shifts (> 0.1 – 0.5% of the baseline frequency) do not appear until a significant crack has already formed and propagated to a measurable length (Andreus & Baragatti, 2009; Česnik et al., 2012; Lorenzino & Navarro, 2015). This is consistent with the mechanism established in Section 2.4.1: early-life frequency changes are typically within measurement uncertainty, not because the underlying physics does not produce them, but because crack areas must grow to a size at which the associated stiffness change exceeds the frequency resolution of the measurement system. This late onset restricts the prognostic horizon available from frequency shift alone.

Modal damping ratio.

The modal damping ratio ζ_n is estimated from the width of the Nyquist circle in the complex FRF plane near each resonance (Česnik & Slavič, 2014; Mei et al., 2019). Its underlying mechanism, crack-face friction, thermoelastic dissipation, and dislocation drag active from the earliest stages of microcrack formation, is established in Section 2.4.1; this ordering, damping preceding detectable frequency shift, has been reported in several experimental studies of metallic specimens (Andreus & Baragatti, 2009; Česnik et al., 2012; Liang et al., 2024).

The practical difficulty with damping as a monitoring feature is its estimation uncertainty. Nyquist circle fitting amplifies noise in the imaginary part of the FRF, and damping estimates carry significantly larger coefficients of variation than frequency estimates from the same measurement. Environmental temperature variations and boundary condition changes also affect damping more strongly than frequency, making baseline correction more critical. Reported magnitudes vary correspondingly: some experimental studies report damping increases of almost 100% between the pristine and near-failure states in metallic specimens (Česnik et al., 2012), while others observe much smaller and less monotonic changes, particularly in early-stage damage (Andreus & Baragatti, 2009). The consequence is a trade-off that recurs throughout the literature: damping is more sensitive but less reliable than frequency shift, and whether it genuinely provides an earlier detection over a statistically confident threshold depends on the specific experimental conditions (Liang et al., 2024; Mei et al., 2019).

Resonance amplitude and compliance.

The FRF peak magnitude $|H(f_n)|$, often expressed as a compliance (mobility at resonance), combines the two mechanisms established in Section 2.4.1: as stiffness decreases and damping rises, the peak both shifts and drops. The fractional compliance change is directly related to the structural flexibility increase caused by the growing crack, making it a physically grounded scalar summary of the overall damage state. Its main limitation as a monitoring feature is sensitivity to excitation level: normalisation against the applied force or shaker acceleration is required to distinguish genuine compliance changes from drift in the actuation chain.

Phase deviation.

As established in Section 2.4.1, phase deviation is expected to be a comparatively sensitive damage signature because the phase slope is steepest at resonance. Magi, Di Maio and Sever (Di Maio & Magi, 2015; Di Maio et al., 2018; Magi et al., 2016a, 2016b) provided the empirical evidence for this: the phase deviation $\Delta\phi = \phi(t) - \phi_0$ constitutes a sensitive and monotonic damage indicator that becomes detectable earlier than frequency shift in fatigue tests on metallic and composite specimens. Subsequent work from the same research lineage has continued to refine the technique, most recently for composite specimens using an interrupted-test framework (Lasen Andrade et al., 2025; Voudouris et al., 2020); phase-deviation monitoring of this kind remains, at the time of writing, a comparatively young technique developed and validated almost exclusively within this one line of work, a point returned to in Section 2.10. Critically for the fixed-frequency protocol used in this thesis, phase deviation is *maximally sensitive* precisely at the operating point: because the phase slope $d\phi/df$ is steepest at resonance (Equation 2.9), any downward drift in f_n produces a disproportionately large phase shift, amplifying the damage signal without requiring any change to the measurement procedure.

Harmonic Distortion Index.

The Harmonic Distortion Index (HDI), whose crack-breathing mechanism and formal definition (Equation 2.13) are established in Section 2.4.2, is a direct indicator of crack-face contact that is largely immune to the stiffness-reduction confounds affecting the linear indicators above (Andreas & Baragatti, 2012; Worden et al., 2002; Xu et al., 2018). Its applicability as a monitoring feature is therefore concentrated in the crack-initiation stage, complementing the linear indicators that dominate the propagation stage (Section 2.4.3). The practical trade-off is measurement sensitivity: super-harmonic components are typically 20–40 dB below the fundamental, near the noise floor of conventional sensing systems, so resolving HDI requires a low noise floor across a wide frequency range and sets a lower bound on the usable vibration amplitude.

2.5.2. Spatial Indicators: Mode-Shape-Derived Features

Global indicators reduce the distributed physical phenomenon of fatigue damage to a scalar at each measurement epoch. While this is sufficient for detecting and tracking damage progression, it discards spatial information about where damage is located and how it distorts the structural mode shapes. Spatial indicators recover this information by exploiting multi-point velocity measurements from a scanning LDV, which provides a discretely sampled mode shape vector $\phi(t) \in \mathbb{R}^N$ at each fatigue block.

Modal Assurance Criterion (MAC).

The MAC quantifies the vector correlation between the current mode shape $\phi(t)$ and the baseline shape ϕ_0 (Allemang & Brown, 1982):

$$\text{MAC}(t) = \frac{|\phi(t)^\top \phi_0|^2}{|\phi(t)|^2 |\phi_0|^2} \quad (2.14)$$

$\text{MAC} = 1$ for shapes that are identical up to a scalar multiple. $\text{MAC} < 1$ indicates shape distortion. For a fatigue crack in a beam, the crack introduces a local stiffness reduction that distorts the mode shape curvature near the crack site, producing a MAC drop even when the natural frequency shift is still within noise (Liang et al., 2024; Montalvão et al., 2006; Singh & Sehgal, 2022). Experimentally, MAC drops of 0.001–0.01 have been reported for early-stage damage in metallic beams, small in absolute terms but potentially detectable with consistent multi-point measurements. The advantage of MAC over global scalar indicators is its insensitivity to amplitude scaling: because Equation 2.14 is a normalised correlation, common-mode variations in response amplitude (e.g., from shaker drift) cancel out. Its limitation is sensitivity to measurement noise at individual scan points, which can produce apparent shape distortions unrelated to damage.

Curvature Damage Index (CDI).

Modal curvature $\kappa(x) = d^2\phi/dx^2$ is proportional to bending strain in an Euler-Bernoulli beam, and is therefore directly related to the local stress state. A fatigue crack produces a local increase in curvature relative to the undamaged pattern at the crack site, because the effective bending stiffness EI is reduced there. The Curvature Damage Index is defined as the spatial field of absolute curvature differences:

$$CDI(x, t) = |\kappa(x, t) - \kappa_0(x)| \quad (2.15)$$

and its spatial maximum, $\max_x CDI(x, t)$, is used as a scalar damage feature. The full spatial distribution of Equation 2.15 can additionally be used to localise the damage: the site where CDI peaks corresponds to the crack location (Liang et al., 2024; Singh & Sehgal, 2022). Curvature estimation from discretely sampled displacement mode shapes amplifies noise through the second-derivative operation, making a sufficient spatial density of scan points a practical prerequisite.

Modal Strain Energy redistribution.

The Modal Strain Energy (MSE) in element j of the beam is proportional to the local curvature squared, $MSE_j(t) \propto \kappa^2(x_j, t)$. Normalised across all elements, the fractional MSE forms a spatial probability distribution over the beam length. Damage causes energy redistribution toward the crack site. The total variation distance between the current and baseline MSE distributions,

$$MSE\text{-}TVD(t) = \frac{1}{2} \sum_j |MSE_j(t) - MSE_{j,0}| \quad (2.16)$$

integrates spatial damage evidence across the full beam length, making it more robust than a single-point curvature metric to noise at individual scan positions (Liang et al., 2024; Singh & Sehgal, 2022). Like the CDI, the practical upper bound of Equation 2.16 is set by the noise amplification inherent in curvature estimation.

Spatial phase standard deviation.

The phase-deviation indicator described above can be extended spatially: instead of computing the phase deviation at a single measurement point, the standard deviation of $\Delta\phi$ across all N scan positions is computed at each fatigue block,

$$\sigma_\phi(t) = \text{std}_{i=1,\dots,N} (\Delta\phi_i(t)). \quad (2.17)$$

In an undamaged beam with a real, spatially coherent mode shape, all scan points share approximately the same phase relationship to the excitation force, and Equation 2.17 evaluates to near zero. A fatigue crack introduces a localised phase anomaly, a spatial discontinuity in the phase field near the crack, which increases the spatial scatter. This indicator is a natural spatial extension of the single-point phase-deviation concept established by Di Maio and collaborators (Di Maio et al., 2018; Magi et al., 2016b), adopted in this thesis's own analysis pipeline rather than reported as an established indicator elsewhere in the literature. It may be more sensitive to localised early-stage damage than the global phase-deviation scalar, which averages the effect over the entire structure, but this expectation has not, to the author's knowledge, been previously tested experimentally.

2.6. Modal Analysis Techniques

The damage indicators described in Section 2.5 are not directly observable quantities: they must be estimated from measured signals through a chain of signal processing and system identification steps. The quality of the damage indicator time series is therefore determined as much by the modal analysis methodology as by the underlying physics. Modal analysis proceeds in three stages once the FRF itself has been estimated (Section 2.2.2): identification of the modal parameters (poles), assembly of

the spatial mode shape from multi-point data, and validation of the resulting modal model before it is trusted as input to later fatigue-monitoring layers. This section describes each stage at the level of the existing literature.

The purpose here is to establish the methodological foundations at the level of the existing literature. The specific implementation choices for the fatigue monitoring pipeline of this thesis are documented in Chapter 3.

2.6.1. Modal Parameter Identification

Two complementary identification strategies are used to extract modal parameters from the FRF. *SDOF local identification* exploits the fact that, near an isolated resonance, the receptance FRF of a viscously damped structure traces a circle in the complex (Nyquist) plane. Fitting this circle (the classical method of Kennedy and Pancu, systematised by Ewins (Ewins, 2000)) yields the natural frequency, damping ratio, and modal residue at that resonance from a narrow frequency band alone. Since it is local and processes one mode at a time, the circle-fit method is fast, robust to noise outside the resonance region, and unaffected by close-mode interference for the well-separated modes typical of simple beam specimens (Allemang & Avitabile, 2022; Ewins, 2000). Its main limitation is the assumption of mode isolation, which breaks down when two resonances lie within a few half-power bandwidths of each other.

When multiple modes must instead be identified together from a single broadband FRF, as in a pre-fatigue baseline characterisation, global *MDOF identification* methods that fit a parametric model across the full frequency range offer better efficiency and consistency than repeated SDOF fits. Two families of MDOF method are in common use.

The Least-Squares Complex Frequency (LSCF, or PolyMAX) method is the current industrial standard in the frequency domain (Allemang & Avitabile, 2022): it fits a matrix-fraction model across a sweep of model orders and distinguishes genuine structural poles from spurious noise-driven ones via a stabilisation diagram, in which physical poles recur consistently across model orders while mathematical artefacts scatter.

An alternative, time-domain family is offered by subspace state-space identification, most widely known through the N4SID algorithm (Van Overschee & De Moor, 1994): rather than fitting a matrix-fraction model in the frequency domain and reading poles off a stabilisation diagram, N4SID estimates a discrete-time state-space model directly from the time-domain input-output data via projections and singular value decomposition, with the system order chosen from the decay of the singular value spectrum rather than from visual inspection of pole clustering across model orders. Frequency-domain and time-domain MDOF methods each have contexts in which they are preferred; a short-duration measurement record, for instance, limits the frequency resolution available to a frequency-domain fit in a way that a time-domain method operating directly on the raw samples does not share.

2.6.2. Mode Shape Assembly from Multi-Point Measurements

Single-point FRF measurement yields the modal parameters at the reference degree of freedom (DOF), but provides no spatial information about mode shapes. Recovering the mode shape vector $\phi_r \in \mathbb{R}^N$ for mode r requires FRF measurements at N spatially distributed points, from which the complex residue at each location is extracted and assembled into the shape vector.

The standard approach is the roving response method: the excitation is fixed at a reference DOF while the response is measured sequentially at each of the N scan positions (Allemang & Avitabile, 2022; Ewins, 2000). For each scan position x_i , the FRF $H(x_i, \omega)$ is computed and the residue $A_r(x_i)$ at the r -th pole is extracted by fixing the poles from the prior MDOF identification (Section 2.6.1) and solving a linear least-squares problem for the residues alone. The assembled vector $\{A_r(x_1), A_r(x_2), \dots, A_r(x_N)\}$ is the mass-normalised mode shape (subject to a common scaling factor), suitable for direct computation of MAC, modal curvature, and MSE-TVD (Ewins, 2000; Magi et al., 2016b).

The spatial resolution of the mode shape, and therefore the sensitivity of curvature and MSE-based damage indicators, depends on the number and distribution of scan points. For curvature estimation via finite differences, the scan spacing must be sufficiently fine to resolve the local curvature anomaly

expected from a fatigue crack: too coarse a grid produces an aliased curvature field that may entirely miss the damage site (Liang et al., 2024; Magi et al., 2016b).

2.6.3. Validation of the Identified Modal Model

Identification, whether SDOF or MDOF, is an estimation procedure and can fail silently: two closely spaced physical modes can be merged into one fitted pole, a genuinely spurious, noise-driven pole can be mistaken for a structural one, or a poorly conditioned residue extraction can distort an assembled mode shape without producing an obviously wrong-looking result. In a fatigue-monitoring context this risk is compounded by the fact that the modal parameters are simultaneously expected to drift for a genuine physical reason, damage accumulation, so an identification artefact and a real damage-driven change must not be confused. The literature offers several standard checks for this.

The stabilisation diagram introduced in Section 2.6.1 is itself a first validation tool: consistency of a candidate pole's frequency and damping across a sweep of model orders is evidence that it is physical rather than a numerical artefact of a particular model order choice. A second, complementary check applies the Modal Assurance Criterion of Section 2.5 to the identified mode shapes themselves rather than to a damage/baseline comparison, an application generally referred to as AutoMAC (Allemang, 2003): computing MAC between every pair of identified modes should yield a near-diagonal matrix, since distinct physical modes are expected to be nearly orthogonal, while a high off-diagonal value signals that the same physical mode has likely been fitted twice at nearby frequencies rather than genuinely resolved as two separate modes. This is the same equation used for damage tracking in Section 2.5, but here the comparison is between two identified modes at a single instant, not between a current and baseline shape of the same mode over time; the two uses should not be conflated.

A further check is the Modal Phase Collinearity (MPC), which quantifies whether an identified mode shape is close to real-valued, as expected for a well-identified, lightly-damped structural mode, or complex, with scan points out of phase with one another (Allemang & Avitabile, 2022; Ewins, 2000). A low MPC value can indicate non-proportional damping in the physical structure, but can equally well indicate measurement noise or a poorly conditioned identification, and is generally interpreted as a data-quality flag alongside coherence (Section 2.2.2) rather than in isolation. Finally, an FRF synthesis check, reconstructing the FRF from the identified poles and residues and comparing it against the originally measured FRF, provides a direct, model-level goodness-of-fit measure that complements the pole-level checks above.

For a single, short modal test these checks are typically applied once, interactively, by an analyst reviewing a stabilisation diagram or a MAC matrix before accepting a set of identified modes. A fatigue-monitoring campaign that identifies modal parameters automatically at every block across a test that may run for days introduces a different challenge: an analyst cannot realistically review every block by hand, yet an unreviewed identification failure partway through a long campaign is difficult to distinguish, after the fact, from a genuine damage-driven change. How this thesis's own pipeline addresses that tension is described in Chapter 4.

2.7. Vibration Measurement for Fatigue Monitoring

The modal analysis techniques described in Section 2.6 place specific and demanding requirements on the measurement system. For the indicators developed in Sections 2.4 and 2.5 to be useful in practice, the sensor must resolve velocity amplitudes at the micrometre level, maintain stable measurement characteristics over fatigue campaigns, operate without contacting the specimen (to avoid mass loading effects on the very modal parameters being tracked), and provide sufficient spatial coverage to assemble mode shape vectors. These requirements, taken together, point unambiguously to non-contact optical interferometry, and specifically to the Laser Doppler Vibrometer (LDV), as the instrument of choice for laboratory resonance fatigue monitoring (Magi et al., 2016b; Rothberg et al., 2017).

This section describes the operating principles and performance characteristics of the LDV, including its fundamental noise limits, and reviews its established role in vibration-based fatigue testing. The section concludes by discussing the scanning LDV configuration that enables spatial mode shape acquisition alongside the scalar modal parameters tracked throughout the fatigue life.

2.7.1. Context: Vibration Sensing Technologies

Vibration sensors are broadly classified by the physical quantity they measure and by whether they contact the structure under test (Kuttner & Rohnen, 2023). Accelerometers (piezoelectric or MEMS devices that measure $a(t)$) are the most widely deployed class due to their ruggedness, low cost, and ease of integration, but they attach physically to the specimen. In a fatigue test on a thin metallic beam, even a lightweight accelerometer (of order 0.5–2 g) shifts the natural frequency by a measurable amount relative to the undamaged baseline, introducing a systematic offset into every frequency-based damage indicator. This mass-loading effect is irreversible over the course of the test and cannot be corrected without knowledge of the modal mass, which is itself uncertain (Allemang & Avitabile, 2022; Ewins, 2000). Strain gauges measure local strain directly and are excellent for stress monitoring at a known point, but they too introduce local stiffness changes and, critically, are themselves susceptible to fatigue failure before the specimen, as demonstrated by Česnik et al. (Česnik et al., 2012).

Non-contact optical techniques avoid these limitations entirely. Among them, LDV has achieved the status of the de facto standard for high-resolution vibration measurement in experimental mechanics and fatigue testing, owing to its combination of nanometre-level displacement sensitivity, zero mass loading, wide bandwidth, and well-understood noise characteristics (Rothberg et al., 2017; Tomasini & Castellini, 2020). These properties make LDV uniquely suitable for the concurrent role required in this thesis: it serves simultaneously as the primary modal characterisation instrument and as the fatigue damage monitoring sensor, without any interference between these two functions.

An emerging alternative to interferometric LDV is quantum photonic vibrometry (QPV), which uses non-classical, single-photon light sources to recover surface velocity without the heterodyne interferometry described in Section 2.7.2 (Rehain et al., 2021). Early demonstrations report sensitivity and bandwidth performance competitive with conventional LDV, and current work extends the approach toward material characterisation applications beyond vibration measurement alone (Huang et al., 2023; Ramanathan et al., 2025), with at least one system already available as a commercial instrument (“Vibrometer - Quantum Computing Inc”, 2018). This technology motivated the original scope of the present project, described in Section 1.3; its unavailability during the experimental campaign led to the adoption of the scanning LDV system used throughout this thesis instead, and a direct QPV-versus-LDV comparison is proposed as future work in Chapter 7.

2.7.2. Physical Principles of LDV

Laser Doppler vibrometry exploits the Doppler frequency shift experienced by coherent light reflected from a moving surface. When a monochromatic laser beam of wavelength λ impinges on a surface vibrating with instantaneous velocity $v(t)$ along the beam axis, the reflected light undergoes a frequency shift:

$$\Delta f(t) = \frac{2 v(t)}{\lambda} \quad (2.18)$$

where the factor of two accounts for the round-trip optical path (Tomasini & Castellini, 2020). For a typical helium-neon or diode laser with $\lambda \approx 633$ nm and a surface velocity of 1 mm/s, Equation 2.18 gives a shift of approximately 3.2 kHz, far below the optical carrier frequency but well within the bandwidth of electronic demodulation circuits.

This frequency shift is recovered interferometrically. In the heterodyne configuration universally preferred in modern LDV systems, shown in Figure 2.3, the incoming laser beam is split into a measurement arm (directed at the target), and a reference arm that passes through an acousto-optic modulator (AOM) introducing a fixed carrier frequency shift f_c (typically 40 MHz). The Doppler-shifted return beam and the frequency-shifted reference beam are recombined at a photodetector, producing a photocurrent:

$$i(t) = \mathcal{R} P_{\text{opt}}(t) \quad (2.19)$$

where $\mathcal{R}$ is the detector responsivity (A/W) and $P_{\text{opt}}(t)$ is the instantaneous optical power incident on the detector (Tomasini & Castellini, 2020). The photocurrent of Equation 2.19 oscillates at the carrier

frequency f_c modulated by the Doppler shift, and demodulation of this signal yields $v(t)$ directly. Heterodyne detection is preferred over the simpler homodyne configuration because it eliminates directional ambiguity (the sign of $v(t)$ is preserved), improves low-frequency stability, and maintains measurement quality in the presence of environmental disturbances (Allemang & Avitabile, 2022; Rothberg et al., 2017).

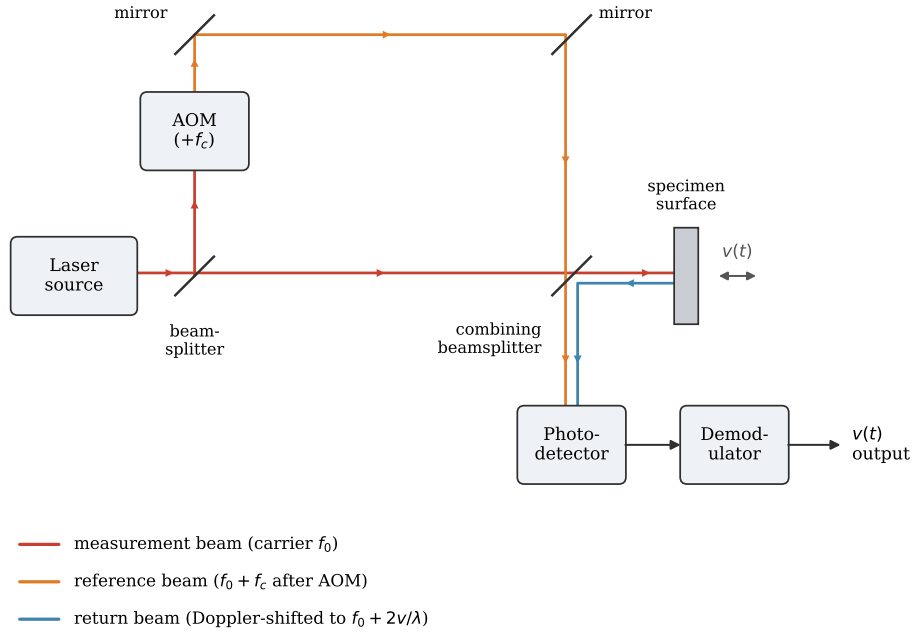


Figure 2.3: Heterodyne LDV configuration. The laser beam is split into a reference arm, shifted by a fixed carrier frequency f_c via an acousto-optic modulator, and a measurement arm directed at the moving specimen surface. The Doppler-shifted return beam recombines with the reference beam at the photodetector; demodulating the resulting photocurrent around f_c recovers the surface velocity $v(t)$.

Displacement $u(t)$ and acceleration $a(t)$ are recovered from $v(t)$ by numerical integration and differentiation, respectively, making the LDV a versatile sensor capable of supplying all three kinematic quantities from a single optical measurement point.

2.7.3. Noise Sources and Sensitivity Limits

LDV measurement quality is governed by a hierarchy of noise sources of decreasing fundamentality (Allemang & Avitabile, 2022; Tomasini & Castellini, 2020). At the bottom sits *shot noise*, the quantum-mechanical floor arising from the discrete, Poisson-distributed arrival of photons at the detector. A well-designed heterodyne instrument is operated in this shot-noise-limited regime, giving the lowest achievable velocity noise floor the hardware can offer (Rothberg et al., 2017; Tomasini & Castellini, 2020). Above this sits digital quantisation noise from the velocity decoder, which in modern instruments is small enough to be inconsequential in practice.

In laboratory conditions, however, the noise floor that actually limits measurement is rarely the fundamental one. *Speckle-induced phase noise*, arising from coherent interference of light scattered from the optically rough specimen surface, together with air turbulence and mounting vibration, typically dominates (Rothberg et al., 2017). Unlike the shot-noise floor, these contributions are not fixed by physics and can be substantially reduced through experimental practice, most commonly by applying retro-reflective tape at the measurement point to stabilise the speckle pattern, and by enclosing the optical path and isolating the instrument from shaker vibration. The practical consequence for fatigue monitoring is that the achievable detection sensitivity is set less by the instrument's data sheet than by how carefully these environmental and surface effects are controlled – a question that depends on the specific instrument and laboratory setup, and is therefore addressed empirically for the LDV unit and

test rig used in this thesis in Chapter 3.

2.7.4. Scanning LDV for Mode Shape Acquisition

A single-point LDV provides a time series at one location on the specimen surface, sufficient for extracting scalar modal parameters (f_n , ζ , phase, HDI) at the reference measurement point. To assemble the full mode shape vector required for the spatial damage indicators of Section 2.5, the measurement must be repeated at N spatially distributed points. This is the function of the scanning LDV (SLDV): a motorised mirror deflects the laser beam to each programmed grid point in sequence, and the FRF at each position is acquired autonomously and stored for subsequent mode shape assembly (Di Maio et al., 2018; Magi et al., 2016b; Tomasini & Castellini, 2020).

The SLDV configuration introduces one complication relative to single-point measurement: because the beam position changes between scan points, the reference for phase comparisons must be maintained consistently across the scan grid. This is achieved either by recording a fixed reference accelerometer signal simultaneously at every scan position, or by using the first scan point as a phase reference and cross-correlating all subsequent points to it. The former approach is preferred because it preserves the force normalisation required for absolute modal mass estimation and residue-based mode shape scaling (Ewins, 2000; Magi et al., 2016b).

In the fatigue monitoring context, the SLDV enables a qualitative capability step beyond single-point measurement: rather than observing that the structure has changed (from scalar indicator drift), it becomes possible to observe where it has changed and how the mode shapes are being distorted. Magi et al. (Magi et al., 2016b) demonstrated that scanning LDV measurements on composite specimens under resonance fatigue loading could detect mode shape distortions associated with delamination initiation before any measurable frequency shift, using a scan grid of 15–20 points along the specimen length. Di Maio et al. (Di Maio et al., 2018) extended this approach to phase-difference indicators computed from simultaneous multi-point velocity and force measurements, showing that the spatial distribution of phase deviation amplifies the damage signal and improves localisation. These demonstrations provide the methodological precedent for the scanning LDV campaign of this thesis, adapted here to metallic specimens under the fixed-frequency resonance protocol described in Section 2.3.

The practical trade-off of SLDV is acquisition time: a full spatial scan of N points requires N sequential measurements, each of sufficient duration to achieve adequate spectral resolution and averaging. For a 30-point scan with 60 s per position, the total scan time is 30 minutes - maybe feasible for a block-cycle protocol where modal characterisation is performed every M fatigue blocks, but too slow for continuous monitoring. This motivates the two-rate measurement strategy in which global scalar indicators (updated every block from the drive-point LDV) are supplemented by less frequent spatial scans (every M blocks), trading temporal resolution for spatial resolution according to the phase of the fatigue life (Di Maio et al., 2018; Magi et al., 2016b).

2.7.5. Optical Fractography as an Independent Ground Truth

Every indicator discussed so far is an inference from vibration response: a frequency shift, a phase deviation, or a mode shape distortion is interpreted as evidence of an underlying crack, but the crack itself is never directly observed while the test is running. Optical fractography, the microscopic examination of a specimen's surface and, once it has separated, its fracture surface, provides the complementary, direct observation against which these vibration-based inferences can be checked. The technique is long established in fatigue research: post-mortem examination of fracture surfaces under optical or scanning electron microscopy is the standard means of identifying crack initiation sites, distinguishing stable propagation from final unstable fracture, and, in favourable cases, reading the striation spacing left by each loading cycle as a direct record of local crack growth rate (Schijve, 2008). Periodic optical imaging of the specimen surface during a fatigue test, rather than only after fracture, extends this to an in-test crack-length record: surface replication techniques and confocal microscopy at scheduled load interruptions have both been used to track crack length and surface crack density as fatigue progresses, providing a ground-truth trajectory against which model predictions of crack growth can be directly compared (Jones et al., 2023; Schijve, 2008).

It is important to be clear about what this technique offers and what it does not. Fractographic interpreta-

tion is largely qualitative: identifying an initiation site, judging whether a propagation front is consistent with a given loading history, and comparing the general character of two fracture surfaces all rely on examiner judgement and domain experience rather than an automated measurement, and only under specialised conditions (clear striation banding, known loading history) does fractography yield a directly quantitative crack-growth-rate curve (Jones et al., 2023).

2.8. Data-Driven Damage Detection and Characterisation

The physical indicators and measurement techniques described in Sections 2.5–2.7 yield, at each fatigue block, a feature vector of modal parameter estimates. Turning this growing time series into actionable information is a well-established problem in the wider structural health monitoring literature, classically organised into a hierarchy of increasingly demanding questions: detecting that damage is present, localising where it is, quantifying its severity, and finally predicting the remaining service life (Rytter, 1993). This section addresses the first three of these questions together, since for the modal indicators considered here they draw on closely related statistical machinery: establishing that a specimen has departed from its healthy baseline and characterising how far it has progressed toward failure are both, at heart, questions about how a growing feature vector compares to a reference distribution, while the spatial indicators of Section 2.5 additionally supply the localisation information the second level of this hierarchy calls for. The fourth and most demanding question, predicting remaining life, is treated separately in Section 2.9, both because it is a conceptually distinct problem, a continuous forecast rather than a comparison against a reference state, and because, as that section will show, it remains comparatively underdeveloped for the present application.

A distinction worth keeping explicit throughout what follows is the one between *detection* and *characterisation*. Detection asks a binary question: does the current feature vector look statistically different from the healthy baseline? Characterisation goes further, asking a continuous question: how far has the structure progressed toward failure? Detection is addressed first below, followed by the sensitivity analysis and index-construction methods that support characterisation.

2.8.1. Baseline Modelling and Multivariate Anomaly Detection

The most widely adopted data-driven approach to damage detection in structural health monitoring is the baseline model paradigm: a statistical representation of the healthy structural state is constructed from measurements acquired before damage accumulation, and subsequent measurements are compared against this model to identify statistically significant deviations (Eltouny et al., 2023; Jolliffe & Cadima, 2016). This is an unsupervised approach: it requires no labelled damage examples and can, in principle, detect any damage mode that perturbs the feature distribution, including modes not anticipated at the design stage.

For a multivariate feature vector $\mathbf{x} \in \mathbb{R}^p$, the natural baseline model is the multivariate Gaussian $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ estimated from the pre-damage measurements. The canonical anomaly score derived from this model is the *Hotelling T^2 statistic*:

$$T^2 = (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \quad (2.20)$$

which is the squared Mahalanobis distance from the healthy cluster centre. The detection threshold is derived from the F -distribution at a chosen false alarm rate, providing a statistically principled criterion (Eltouny et al., 2023). The T^2 statistic of Equation 2.20 integrates information from all features simultaneously, making it more sensitive to subtle multi-feature deviations than any single univariate threshold, while remaining interpretable in terms of the underlying multivariate Gaussian model.

For detecting gradual, progressive damage, instantaneous threshold tests are complemented by the *Cumulative Sum* (CUSUM) control chart (Page, 1954). CUSUM accumulates small, consistent upward drifts in the anomaly score across successive blocks, triggering an alarm when the cumulative sum exceeds a threshold h :

$$S(t) = \max(0, S(t-1) + s(t) - k) \quad (2.21)$$

where $s(t)$ is the current anomaly score, k is the allowance parameter (typically set to half the expected shift magnitude), and the chart signals when $S(t) > h$. The recursion of Equation 2.21 is particularly well-matched to fatigue damage because structural degradation is inherently gradual: a small but consistent upward trend in T^2 across many blocks will reliably trigger CUSUM well before any individual block exceeds its instantaneous threshold, a behaviour also exploited by Mishra et al. (Mishra et al., 2015) in a multivariate CUSUM chart applied directly to sensor data from fatigue-loaded composite specimens.

A complementary non-parametric detector is the *Isolation Forest* (Liu et al., 2008), which makes no Gaussian assumption and detects anomalies by measuring how easily a new observation can be isolated by random axis-aligned splits of the feature space. Anomalous points (those far from the dense healthy cluster) isolate faster and receive lower anomaly scores. The Isolation Forest is particularly useful when feature distributions are non-Gaussian or when a small number of highly sensitive features dominate, since it adapts to the geometry of the data rather than assuming ellipsoidal clusters. Toufigh and Ranjbar (Toufigh & Ranjbar, 2024) combined it with a deep autoencoder for exactly this purpose in an unsupervised structural damage detection framework.

Where researchers broadly agree is that combining instantaneous and sequential detectors (such as T^2 and CUSUM) with a non-parametric complement (Isolation Forest) provides more robust and earlier detection than any single method (Elouny et al., 2023). The principal open question is not which method is best in general, but which combination provides the best detection performance for the specific feature set and damage rate of a given experimental context.

2.8.2. Dimensionality Reduction and Indicator Sensitivity Ranking

The multi-indicator feature vector defined in Section 2.5 is intentionally redundant: it contains both fast-responding and late-responding indicators, both linear and nonlinear features, and both global and spatial quantities. Understanding which features carry the most damage-relevant information, and at what stage of the fatigue life they are most informative, requires systematic sensitivity analysis.

Principal Component Analysis (PCA) provides a first decomposition of the feature matrix $\mathbf{X} \in \mathbb{R}^{N \times p}$ (blocks $\times$ features) into orthogonal directions of maximum variance (Jolliffe & Cadima, 2016). When a single progressive damage process dominates the experiment, the first principal component (PC1) captures the dominant direction of feature co-variation and its score time series $z_1(t) = \mathbf{X}\mathbf{v}_1$ constitutes a data-driven composite damage signal that is more robust to individual noisy features than any single indicator. The fraction of total variance explained by PC1 is a direct measure of how coherently the feature set tracks a single underlying damage process: values above 0.6 suggest that a single latent variable (damage state) explains most of the observed variability (Bandara et al., 2014; Jolliffe & Cadima, 2016).

Beyond PCA, supervised sensitivity ranking provides a more targeted answer to which features are most predictive of the remaining life fraction rather than merely most variable. The framework used in this thesis employs gradient-boosted regression trees trained to predict the normalised cycle fraction $\tau = N/N_f$ from the feature vector, with SHAP (SHapley Additive exPlanations) values used to decompose each prediction into per-feature contributions (Lundberg & Lee, 2017). SHAP values satisfy the efficiency, symmetry, dummy, and linearity axioms from cooperative game theory, making them the theoretically principled solution to the feature attribution problem.

2.8.3. Damage Index Construction

Anomaly detection establishes when the structure has departed from its healthy baseline. A damage index $D(t) \in [0, 1]$ goes further: it characterises how far the structure has progressed toward failure, normalising the damage state to a common scale that enables comparison across specimens, loading conditions, and indicator types. The physical motivation for a bounded, monotonically increasing index comes directly from CDM (Section 2.1): the scalar CDM damage variable D maps the pristine state ($D = 0$) to complete failure ($D = 1$), and its evolution follows a characteristic S-shaped trajectory governed by Paris-law crack growth dynamics.

Constructing such an index from raw sensor features, rather than selecting a single one and using it directly, is itself an established problem in the wider prognostics and health management (PHM)

literature, generally referred to as health indicator construction (Atamuradov et al., 2020). Two broad strategies recur across applications. The first selects or transforms a single physically meaningful feature, for example a frequency shift normalised by its own maximum observed change, and relies on that feature's own established relationship to damage; its weakness is that a single feature may plateau or saturate before failure if the physical quantity it tracks stops changing while damage continues to accumulate by other mechanisms. The second fuses multiple features into one composite index, either through a supervised weighting scheme or an unsupervised projection such as PCA (Section 2.8.2), trading the interpretability of a single physical quantity for robustness to any one feature's idiosyncratic noise or early saturation (Atamuradov et al., 2020).

Whichever strategy produces it, a candidate index still needs some basis for judging whether it is actually any good, since a bounded, superficially S-shaped index can be constructed from data that tracks damage poorly. Coble and Hines (Coble & Hines, 2009) formalised three criteria for this purpose that have since become standard in the PHM literature: *monotonicity*, the degree to which the index trend is consistently increasing rather than noisy or reversing; *trendability*, the degree to which multiple specimens or units produce index trajectories of similar shape once normalised to a common life fraction; and *prognosability*, the degree to which the index reaches a similar value at failure across a population, a precondition for a failure threshold learned from one specimen to be meaningful for judging another. These three criteria provide the common basis on which competing damage index formulations, whichever features or fusion strategy produced them, can be compared.

2.8.4. Learned Health Indicators: Autoencoder-Based Damage Indices

Both strategies described above share a common structure: each is built from features chosen because their physical or statistical relationship to damage is already understood. An alternative paradigm, now well established in the broader structural health monitoring (SHM) literature, dispenses with feature engineering entirely and instead learns a representation of the healthy state directly from raw or near-raw response data. A convolutional or recurrent *autoencoder* is trained to reconstruct measurements drawn only from the healthy (pre-damage) phase of a structure's life. Since the network has only ever learned to reconstruct healthy data, its reconstruction error on subsequent measurements rises when the input departs from the healthy manifold, providing a damage indicator that requires no prior assumption about which spectral features matter (Okur et al., 2025). A complementary indicator can be drawn from the network's compressed latent representation. Rather than measuring how well the input is reconstructed, the deviation of the latent code itself from its healthy-phase distribution is tracked, which can register a structurally meaningful change in input shape even when the full-resolution reconstruction error has not yet moved appreciably, since a single latent dimension can be sensitive to a spectral feature that contributes little to overall reconstruction loss.

This reconstruction-error paradigm has been demonstrated extensively on civil infrastructure, most commonly on bridges under ambient or operational excitation, where autoencoder architectures ranging from simple convolutional networks to attention-based LSTM variants have been shown to detect, and in some cases localise, discrete damage events such as loosened connections or induced support damage (Ghazimoghdam & Hosseinzadeh, 2024; Okur et al., 2025). A recurring finding in this literature is that autoencoder reconstruction error and classical statistical indices such as the Hotelling T^2 or PCA-based scores are not redundant because the autoencoder learns an arbitrary nonlinear manifold of the healthy state rather than assuming a Gaussian or linear-subspace structure, it can be sensitive to forms of deviation that a hand-engineered feature set does not encode, and several studies report explicit comparisons in which a learned reconstruction-error indicator outperforms a traditional PCA baseline on the same data (Li et al., 2023). This motivates treating a learned indicator as a genuinely complementary damage index in the present framework. The expectation, consistent with the wider literature, is that the two families will disagree on the precise point of onset, and that this disagreement is itself informative about which aspects of the FRF a given damage mechanism perturbs.

Two gaps in this literature are directly relevant to the present work. First, the autoencoder-SHM literature to date is concentrated almost entirely on civil structures under ambient, broadband, or operational excitation. Metallic fatigue specimens under controlled, narrowband resonance excitation, where the input to the network is a deliberately well-characterised FRF rather than an uncontrolled ambient response appear not to have been studied with this technique, and the network architectures and training

assumptions developed for the civil case have not been validated against the damage mechanisms (fatigue crack initiation and propagation under cyclic stress) that are the subject of this thesis. Second, where reconstruction-error and hand-engineered indices have been compared, the comparison has generally been framed as a competition to identify the better-performing method, rather than as a joint multi-indicator analysis of the kind already applied to the hand-engineered feature set in this chapter (PCA, SHAP, sensitivity ranking). Applying the same comparative apparatus to a learned indicator, within a single consistent experimental campaign, is a natural extension of the multi-indicator gap identified above and is addressed, on an exploratory basis, alongside the primary hand-engineered analysis in this thesis.

2.9. Remaining Useful Life and Prognosis

Whichever damage index formulation is used, hand-engineered or learned, characterising the current damage state answers only the first three levels of the hierarchy introduced in Section 2.8. The fourth and most demanding question remains open: given a damage index time series $D(\tau)$ as a function of normalised cycle fraction $\tau = N/N_f$, at what future τ_f will D reach the failure threshold D_f , and how uncertain is that estimate given the data available up to the current block? This is the domain of prognosis, and it is, by a wide margin, the least mature of the four levels for the modal-parameter-based indicators considered here, an imbalance developed further as Gap 2 in Section 2.10.

General path models, in which a chosen degradation measure is extrapolated to a failure threshold using a parametric curve fitted to its own observed history, are among the earliest and most widely used data-driven RUL methods (Lu & Meeker, 1993). The most physically motivated parametric model for $D(\tau)$ in the present context is the logistic (S-curve) function:

$$D(\tau; \beta, \tau_{\text{mid}}) = \frac{1}{1 + \exp(-\beta(\tau - \tau_{\text{mid}}))} \quad (2.22)$$

where β controls the steepness of the damage acceleration and τ_{mid} is the cycle fraction at the inflection point. The model of Equation 2.22 is consistent with Paris-law crack growth (S-shaped CDM evolution, Section 2.1), has a natural upper asymptote at $D = 1$ preventing non-physical extrapolations, and can be updated incrementally as new block data arrives. Confidence intervals on the RUL estimate can be propagated from the parameter covariance matrix of the nonlinear least-squares fit, widening early in life (when only the lower half of the S-curve is observed) and narrowing as damage approaches the inflection point.

The principal limitation of a general path model is that it is fitted to a single specimen's own history and carries no information about how other, similar specimens have behaved. For datasets where several run-to-failure histories are available, *Gaussian Process* (GP) regression provides a non-parametric, data-driven complement that incorporates population-level variability directly into the predictive uncertainty (Rasmussen & Williams, 2006). A GP defines a prior distribution over functions $D(\tau)$; conditioned on the observed data from completed tests, its posterior gives a calibrated predictive distribution for the remaining life of a new, in-progress specimen. GP regression has been applied directly to fatigue-life prediction for metallic specimens under variable-amplitude loading (Gao et al., 2022) and under multi-axial stress states (Karolczuk & Słowski, 2022), and entropy-based probabilistic fatigue prognosis has been benchmarked using PHM performance metrics of the kind adopted below (Guan et al., 2009); in each case, however, the specimen populations used for training are substantially larger than what a resonance fatigue campaign on discrete specimens can typically provide, a constraint returned to in the Discussion.

A central question for any RUL method, parametric or GP-based, is not merely whether it eventually converges to an accurate estimate, but how early in a specimen's life a useful prediction becomes available. Saxena et al. (Saxena et al., 2010) formalised this as the *prognostic horizon*: the earliest point in a system's operating life at which a prognostic algorithm's predictions satisfy a chosen accuracy or confidence criterion, and continue to do so for the remainder of life. A method with a longer prognostic horizon provides more actionable warning before failure, all else being equal, which makes prognostic horizon a natural companion metric to the monotonicity, trendability, and prognosability criteria of

Section 2.8.3 when comparing candidate index-and-method combinations. The literature on modal-parameter-based RUL estimation specifically for metallic specimens under resonance fatigue, evaluated against these criteria, is sparse: individual studies have tracked frequency or damping through to failure (Česnik et al., 2012; Wang et al., 2008), but none has systematically evaluated the prognostic horizon achievable from a multi-indicator feature set under controlled resonance fatigue conditions on metals. This is the core gap that Section 2.10 formalises and that this thesis addresses experimentally.

2.10. Research Gap

The preceding sections have mapped the landscape of knowledge on which this thesis builds. Before identifying what is missing, it is useful to state clearly what the literature has established beyond reasonable dispute.

It is well-established that fatigue damage in metallic structures produces progressive and measurable changes in dynamic response: natural frequencies decrease as stiffness is lost (Česnik et al., 2012; Liang et al., 2024), damping ratios increase due to energy dissipation at crack faces (Liang et al., 2024; Mei et al., 2019), and mode shapes are spatially distorted near the crack site (Liang et al., 2024; Montalvão et al., 2006). The physical mechanisms underlying these changes are well understood through CDM and fracture mechanics (Lemaitre & Chaboche, 1990; Murakami, 2012; Schijve, 2008), and resonance vibration fatigue testing has been demonstrated as an effective accelerated methodology for generating and monitoring this damage in controlled laboratory conditions (Česnik et al., 2012; Slavič, 2021). Scanning LDV has been shown to be capable of resolving the mode shape distortions associated with early-stage damage on beam specimens under resonance loading (Di Maio et al., 2018; Magi et al., 2016b), and a range of statistical and data-driven methods can detect departures from the healthy baseline once they cross the measurement noise floor (Eltouny et al., 2023; Jolliffe & Cadima, 2016). A recent review of the broader fatigue-monitoring field reaches an independent, consistent conclusion: research on fatigue calculation, testing, and modelling is extensive, but studies that actually monitor the fatigue process in real time, rather than characterising it before or after the fact, remain comparatively scarce (García-Fernández et al., 2023).

These are the established foundations. Three interlocking gaps remain, which together define the research space occupied by this thesis.

2.10.1. Gap 1: No Systematic Multi-Indicator Comparison Under Fixed-Frequency Resonance Testing on Metals

The individual indicators catalogued in Section 2.5 – natural frequency shift, damping ratio change, resonance amplitude, phase deviation, harmonic distortion index, MAC, modal curvature, MSE redistribution, and spatial phase scatter – have each been studied, but almost invariably in isolation or in small subsets. Studies on metallic specimens have generally tracked only one or two of these indicators at a time (Česnik et al., 2012), while studies that include the spatial indicators have generally been carried out on composite rather than metallic specimens (Di Maio et al., 2018; Magi et al., 2016b). Česnik et al. (Česnik et al., 2012) tracked frequency and damping simultaneously on aluminium specimens under harmonic resonance loading and demonstrated that damping changes precede frequency shifts, but did not extend the analysis to phase-based or spatial indicators, nor did they quantify the relative sensitivity of each indicator as a function of consumed life fraction. Lorenzino and Navarro (Lorenzino & Navarro, 2015) characterised the non-monotonic frequency evolution (hardening then softening) in aluminium but did not study damping, phase, or mode shape indicators in the same campaign. Studies that do include spatial indicators typically operate under different loading conditions (servo-hydraulic testing or base-excitation with frequency tracking) rather than the fixed-frequency resonance protocol (Di Maio et al., 2018; Magi et al., 2016b).

This fragmentation is not simply a matter of no one having gotten around to a systematic comparison; it reflects the narrow base of research programmes the field currently draws on. The general vibration-fatigue-monitoring literature cited throughout this chapter traces overwhelmingly to two sustained research programmes: the Laboratory for Dynamics of Machines and Structures at the University of Ljubljana (Section 2.3), and the phase-deviation lineage originating with Magi, Di Maio and Sever, now continued at the University of Twente (Section 2.4.1). Where even the demonstrations of individ-

ual indicators are concentrated in two labs rather than distributed across a broad field consensus, a systematic multi-indicator comparison of the kind this thesis undertakes has correspondingly little prior work to draw on.

The consequence is that no published study has simultaneously tracked the full set of global and spatial modal indicators through the complete fatigue life of metallic specimens under fixed-frequency harmonic resonance excitation, characterised their relative onset timing and sensitivity as a function of consumed life fraction, and assessed their redundancy and complementarity in a multi-indicator feature space. This is the specific experimental and analytical gap that Research Question R2 addresses.

2.10.2. Gap 2: Absence of a Prognostic Horizon Analysis for Modal-Parameter-Based RUL Estimation on Metals

The detection literature is comparatively mature: numerous methods have been validated for reliably flagging the onset of damage from modal parameter drift (Česnik et al., 2012; Eltouny et al., 2023). The prognosis literature, estimating how much life remains once damage has been detected, is substantially less developed for modal-parameter-based indicators on metallic specimens.

Work on vibration-based fatigue life estimation for metals has largely proceeded along two separate tracks. The first track uses modal parameters to predict fatigue life from material and geometry properties before testing begins, applying spectral methods such as Dirlik's formula to the response PSD (Dirlik & Benasciutti, 2021; Slavič, 2021). This is a design-time calculation, not an in-service monitoring approach. The second track uses frequency shift or compliance change as an empirical damage metric, fitting the measured evolution to a CDM-consistent curve and extrapolating to failure (Česnik et al., 2012; Di Maio et al., 2018). While the feasibility of this approach has been demonstrated on single specimens or small cohorts, no study has: (a) compared the single-feature and fused/composite index-construction strategies of Section 2.8.3 against one another, using the monotonicity, trendability, and prognosability criteria that have become standard in the wider PHM literature, to assess which yields the most consistent and predictable progression curve; (b) characterised the prognostic horizon as a function of the choice of indicator and damage index formulation; or (c) validated the prognosis against an independent dataset of specimens to provide a population-level estimate of prognostic accuracy with uncertainty quantification.

Liang et al. (Liang et al., 2024) addressed the RUL estimation problem for composite laminates using frequency shift as the sole indicator, reporting useful predictions from approximately 40–60% of fatigue life. Equivalent population-level prognostic horizon results have not been reported for metallic specimens under resonance fatigue conditions using a multi-indicator approach, and the comparison between parametric trend extrapolation and data-driven Gaussian Process regression for this specific problem has not been attempted. Research Questions R1 and R3 directly address this gap.

2.10.3. Gap 3: Validating Resonance-Based Testing Against Standard Fatigue Characterisation for Metals

Resonance vibration fatigue testing has been established as a practical, order-of-magnitude faster alternative to conventional constant-amplitude fatigue testing on servo-hydraulic rigs (Section 2.3), and a handful of studies have directly validated the resulting fatigue-life data against standard testing methods (Furman et al., 2025; George et al., 2004; Teng et al., 2022). This validation, however, has generally been carried out under *resonance-tracked* excitation, where the stress amplitude in the fatigue zone is held constant throughout the specimen's life by continuously adjusting the drive frequency to compensate for the shifting resonance (Česnik et al., 2012).

The fixed-frequency protocol adopted in this thesis (Section 2.3.2) does not carry this guarantee: because the excitation frequency is not adjusted as the natural frequency drifts downward under accumulating damage, the specimen progressively detunes from resonance, and the true stress amplitude in the fatigue zone is not held constant across the test even though the applied force amplitude is. Whether, and to what extent, this detuning-induced stress relaxation distorts the resulting fatigue-life data relative to a conventional constant-stress-amplitude test, and therefore whether S-N characterisation performed under a fixed-frequency protocol is directly comparable to standard testing methods (ASTM E466, ISO 1099) or to resonance-tracked results, is a question the existing literature does not

answer directly for metallic specimens.

The trade-off is not without a compensating benefit. Because the excitation does not adapt to the shifting resonance, the detuning itself is directly encoded in the phase and amplitude of the response, amplifying the damage signal for monitoring purposes without requiring any additional instrumentation (Di Maio et al., 2018; Lorenzino & Navarro, 2015). Whether this monitoring advantage is worth the S-N characterisation uncertainty it introduces, and how large that uncertainty actually is, has not been quantified for metallic specimens.

This is the gap Research Question R4 addresses: establishing the S-N characterisation actually achievable under the fixed-frequency protocol used throughout this campaign, and situating it explicitly against both the standard testing literature and the resonance-tracked alternative, so that the monitoring advantage and the load-control accuracy each protocol offers can be weighed against one another with actual data rather than by assumption.

2.10.4. Thesis Positioning and Research Questions

Together, these three gaps define a coherent and practically motivated research space: the use of modal parameter evolution, tracked continuously by a scanning LDV under fixed-frequency resonance excitation, as the primary information source for both damage detection and remaining life prognosis in metallic specimens, with a systematic multi-indicator analysis that provides the empirical foundation for answering when, and how reliably, such prognosis can be achieved.

It should be acknowledged that the research direction of this thesis emerged from a topic change forced by external circumstances: the original plan to benchmark a quantum photonic vibrometer required the instrument to be operational, and its failure prompted a reorientation toward the fatigue monitoring problem that the existing experimental infrastructure (shaker, LDV, data acquisition system, and the analytical pipeline already under development) was already well-positioned to address. Rather than a limitation, this context underscores that the research questions are grounded in genuine experimental capability and that the gap identified above is a natural and scientifically productive target for the available methodology.

The four research questions that structure this thesis map directly onto the identified gaps:

- R1** *How does the vibration response evolve as fatigue damage accumulates under resonant vibration loading?* – Addresses the empirical characterisation of the full indicator set (Gap 1), providing the foundational dataset for R2 and R3.
- R2** *Which vibration-based indicators are most sensitive to early fatigue damage?* – Addresses the comparative sensitivity and onset-timing question (Gap 1), producing the ranked indicator table required to guide future monitoring system design.
- R3** *Can vibration response trends be used to estimate remaining fatigue life?* – Addresses the prognostic horizon analysis and population-level RUL validation (Gap 2), providing the quantitative answer to the central monitoring question.
- R4** *How does resonance-based vibration fatigue testing compare to traditional fatigue tests?* – Addresses the validation of the fixed-frequency resonance protocol against standard fatigue characterisation methods (Gap 3), including its implications for S-N equivalence and damage rate interpretation.

The methodology employed to address these questions is described in Chapter 3.

3

Experimental Design and Development

This chapter describes the experimental design and development work that precedes the analysis pipeline proper: the synthetic prediction toolkit, specimen design, instrumentation, and pre-campaign characterisation, culminating in a comparison between predicted and measured behaviour that justifies the final test protocol. It is procedural and developmental in character rather than results-bearing: none of the outputs presented here directly answer Research Questions R1–R4, though they establish the experimental foundation on which every subsequent chapter’s answers depend. The analysis pipeline built on top of this experimental foundation, comprising two independently constructed indicator tracks that converge on a shared damage-index and remaining-useful-life estimation step, alongside independent S-N characterisation and live-monitoring/validation investigations, is described separately in Chapter 4.

The experimental campaign is organised into three sequential stages, each building on the outputs of the previous one, as summarised in Figure 3.1. The Synthetic Data Toolkit’s predictions (Section 3.1) set the initial expectations for specimen geometry, drive frequency, and excitation level; these are confirmed and, where necessary, refined experimentally at each subsequent stage, arriving at the fatigue test protocol that is then frozen and applied to the two main specimen groups. Both groups feed the analysis pipeline described in Chapter 4. The first stage, pre-campaign characterisation, quantifies the measurement performance of the Laser Doppler Vibrometer (LDV) and establishes the baseline modal properties and linear operating regime of the specimens. This stage encompasses both sensor characterisation tests and a load configuration study in which the excitation amplitude, frequency, and geometric setup were varied across a set of preliminary specimens until a configuration capable of producing consistent, reproducible fatigue fracture at the intended location was identified. The second stage, the fatigue campaign itself, subjects a population of ten nominally identical aluminium beam specimens to a fixed-frequency resonance fatigue protocol, collecting continuous modal parameter data from the healthy baseline onward; eight of these specimens were run to complete fracture under the standard protocol, while two were run under a deliberately interrupted protocol and combined with periodic optical microscopy as an independent validation track (Section 3.2.3). The third stage, described in Chapter 4, is the data analysis pipeline itself.

A note on terminology is useful before proceeding, since it recurs throughout this chapter. The original project, before its reorientation following the QPV instrument failure (Section 1.3), was structured around four sequential phases: Phase 1, sensor characterisation; Phase 2, specimen/sample characterisation; Phase 3, a healthy-baseline resonance run; and Phase 4, the resonance fatigue loading run itself. This thesis retains that numbering rather than renumbering from Phase 1, since the underlying test protocols and, in places, the raw data themselves carry over directly from the original project. Phases 1 and 2 are described in this chapter by their individual test identifiers (T1–T4 and S1–S5 respectively, Section 3.4) rather than by phase number, since no single monolithic “Phase 1” or “Phase

2" test exists; Phases 3 and 4 retain their original numbering directly, since both are still run as single, continuous protocols essentially unchanged from the original design (Section 3.5). The reader should therefore expect "Phase 3" and "Phase 4" throughout this chapter without a corresponding "Phase 1" or "Phase 2" label ever appearing on a specific test – this is intentional, not an omission.

The fixed-frequency resonance protocol employed throughout the fatigue campaign was adopted for the reasons set out in Section 2.3.2 and Section 2.10 of the Literature Review: it trades load-control accuracy for a directly interpretable phase-deviation signal, at the cost of an S-N characterisation whose comparability to standard testing methods this thesis itself investigates (Research Question R4).

Throughout the chapter, the term *block* refers to the fundamental unit of the test protocol: a fixed number of fatigue loading cycles followed by a broadband LDV scan for modal parameter extraction. All time axes in the analysis are expressed in terms of the normalised cycle fraction $\tau = N/N_f$, where N is the cumulative cycle count at the end of a given block, and N_f is the total cycles to fracture for that specimen. This normalisation removes the natural scatter in fatigue life between nominally identical specimens before any cross-specimen comparison is performed.

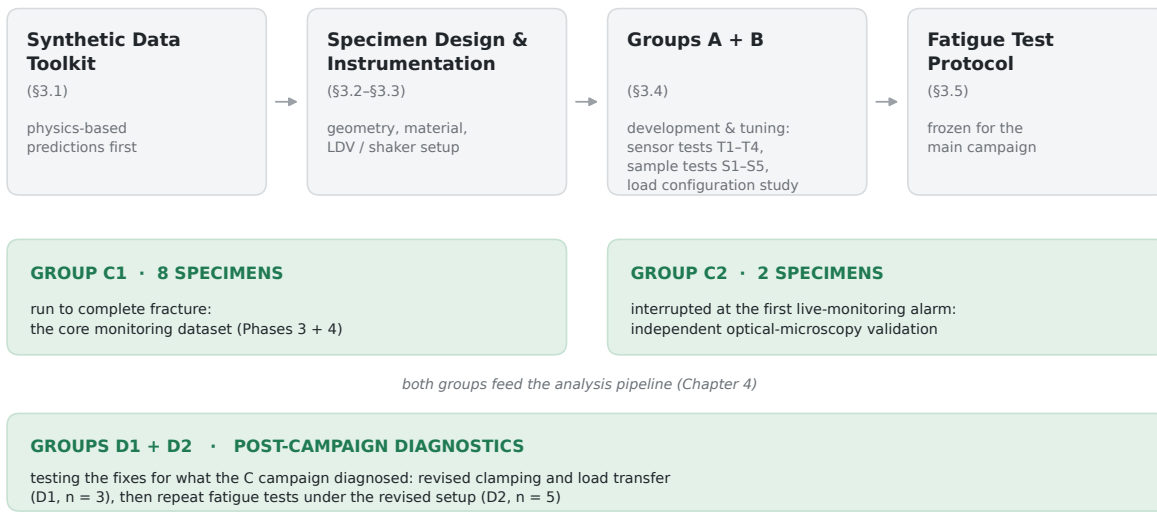


Figure 3.1: Structure of the experimental campaign described in this chapter, with section references. Groups D1 and D2 were added after the main campaign had concluded and are described in Section 3.2.3; their results are reported separately in Section 5.8.

3.1. Synthetic Data Toolkit

Before any experimental hardware was operated, a physics-based simulation environment (referred to throughout this thesis as the Synthetic Data Toolkit) was developed to predict the expected behaviour of the fatigue monitoring system under controlled conditions. The toolkit served two distinct purposes in the overall project. During the design phase, it was used to explore the parameter space of specimen geometry, material properties, and loading conditions, providing quantitative predictions of natural frequency, stress amplitude, expected fatigue life, and the anticipated magnitude of modal parameter shifts at various stages of damage. These predictions informed the selection of the final specimen geometry and the target excitation amplitude before a single fatigue test was run. During and after the experimental campaign, the same toolkit was used as an interpretive reference: simulated damage evolution curves and indicator sensitivity rankings were compared post-hoc against the experimental data to assess where the physics-based model and the measured behaviour agreed, and where discrepancies pointed to phenomena, such as crack closure, localised plasticity, or fixture nonlinearity, not captured by the underlying assumptions. The toolkit is therefore neither a validation target nor a ground truth, being instead a physically grounded interpretive companion to the experimental results.

3.1.1. Physical Model

The toolkit models the fatigue specimen as a rectangular plate or beam with configurable geometry, material, boundary conditions, and harmonic loading. Damage accumulation follows the continuum damage mechanics (CDM) framework and the resulting frequency–damage relationship established in Section 2.1 of the Literature Review; the toolkit implements this directly as a per-element stiffness degradation,

$$E(D) = E_0 (1 - \alpha D) \quad (3.1)$$

with $\alpha = 1.5$ for the aluminium alloys used in this study, consistent with published CDM parameters for this material class (Murakami, 2012).

The number of cycles required to drive each element from its current damage state to failure is computed from the Basquin S-N law introduced in Section 2.1, anchored to the element's local stress state:

$$N_f = \left(\frac{\sigma_{\text{eff}}}{\sigma'_f} \right)^{1/b} \quad (3.2)$$

where σ_{eff} is the effective stress amplitude at the element and $\sigma'_f = \sigma_{\text{end}} \times (10^6)^{-b}$ is derived from the material endurance limit σ_{end} , ensuring the model produces physically meaningful fatigue lives at the stress amplitudes relevant to this study. For Al 6061-T6, the parameters used are $\sigma_{\text{end}} = 97$ MPa and $b = -0.106$ (Table 3.1), consistent with published fatigue design data for this alloy (MatWeb, LLC, 2026).

At each simulation step, the effective stress at every element is computed from the current modal state, the element damage field is updated by integrating the Basquin law of Equation 3.2 for a fixed number of cycles, and the stiffness matrix is recomputed from the updated D field. This iterative loop produces a time series of modal parameters at each damage increment, mimicking the block structure of the experimental protocol described in Section 3.5.

This entire physical model rests on small-deflection, linear elastic theory throughout: the Euler-Bernoulli beam formulation, the CDM stiffness degradation of Equation 3.1, and the modal decomposition underlying all three analysis tiers (Section 3.1.2) each assume a stiffness matrix that is independent of vibration amplitude. This assumption holds well for the low-amplitude broadband scans used for modal identification throughout this thesis, but is not guaranteed to hold at the higher vibration amplitudes used during Phase 4 harmonic fatigue loading, where geometric nonlinearity (amplitude-dependent stiffness from large-deflection effects) becomes measurable for a specimen of this slenderness, as confirmed experimentally by the pre-campaign nonlinearity test S4 (Section 3.4). Toolkit predictions at the harmonic-loading amplitude are therefore better understood as a qualitative guide to expected trends than as a rigorous quantitative forecast; this distinction is revisited throughout this chapter whenever a toolkit prediction is compared against measured behaviour.

3.1.2. Three-Tier Modal Analysis Hierarchy

Modal analysis and stress computation are implemented at three levels of physical fidelity, which the toolkit documentation refers to as Tier 1, Tier 2, and Tier 3. The tier is selected explicitly by the user; there is no automatic fallback between levels.

- **Tier 1** (uniform analytical) applies the Euler–Bernoulli beam formula with a single global effective stiffness E_{eff} , computed as the spatial average of $E(D)$ over all elements. All natural frequencies shift by the same fractional amount, and the stress field is computed from the cantilever bending formula with a Dynamic Amplification Factor (DAF). Tier 1 runs in seconds and is used for rapid parameter sweeps.
- **Tier 2** (spatial Rayleigh-quotient) weights each element's contribution to the stiffness matrix by its local $E(D_e)$, so that different modes respond differently to localised damage depending on the curvature energy they deposit at the damaged region. A mode whose shape has high curvature near the damage site shifts more rapidly than one whose nodal line passes through it. Tier 2

is the preliminary design tool used in this project: it predicts which modes are most sensitive to a given damage location for a given, spatially-varying stiffness field, and therefore informs both the scan grid design and the selection of modes to track. Tier 2's stiffness field is nonetheless a smoothed representation rather than a true discretised geometry: it does not resolve the notch as an explicit geometric cutout the way a meshed model does, so the specific stress concentration a real notch produces is not captured at this tier.

- **Tier 3** (Abaqus FEM) replaces the analytical estimators with a full finite element eigenvalue extraction (Abaqus `FrequencyStep`, Lanczos solver) and a harmonic response analysis (Abaqus `SteadyStateDirect`) that provides element-level von Mises stress fields. Tier 3 is the only tier that resolves the notch as an explicit geometric feature in the mesh, and accordingly serves as the ground-truth reference against which the analytical tiers are cross-checked. It was used selectively during the design phase to validate the Tier 2 predictions for the final specimen geometry. Despite qualitative agreement between Tier 2 and Tier 3, quantitative results differ by more than an order of magnitude, as Figure 3.2 illustrates. Both tiers predict a monotonic decrease in natural frequency over life, but the magnitudes are not comparable. Over the full simulated life the analytical tier predicts a normalised frequency drop of approximately 6% for mode 1, rising to between 18% and 23% for modes 2 to 4, whereas the finite element tier predicts a drop of roughly 91% for every one of the four modes.

Two features of the Tier 3 curves indicate that this is a property of the model rather than a physical prediction to be taken at face value. The four modes fall along essentially the same line, where a localised stiffness loss would be expected to affect each mode in proportion to the curvature it carries at the damage site; and the decline is linear in cycle count rather than accelerating toward failure, which is not the behaviour the underlying continuum damage formulation of Section 3.1.1 produces. A 91% frequency reduction also corresponds to the loss of almost all of the beam's effective stiffness, a state well beyond the damage levels at which the tracked indicators are of interest. This divergence is why Tier 2 is used to rank modes by relative sensitivity rather than to predict absolute frequency shifts (Section 3.1.2), and why the comparison of Section 6.1.6 treats the toolkit prediction as qualitative. This is due to not only the stress concentration effects, but also by the way each step is computed at each tier.

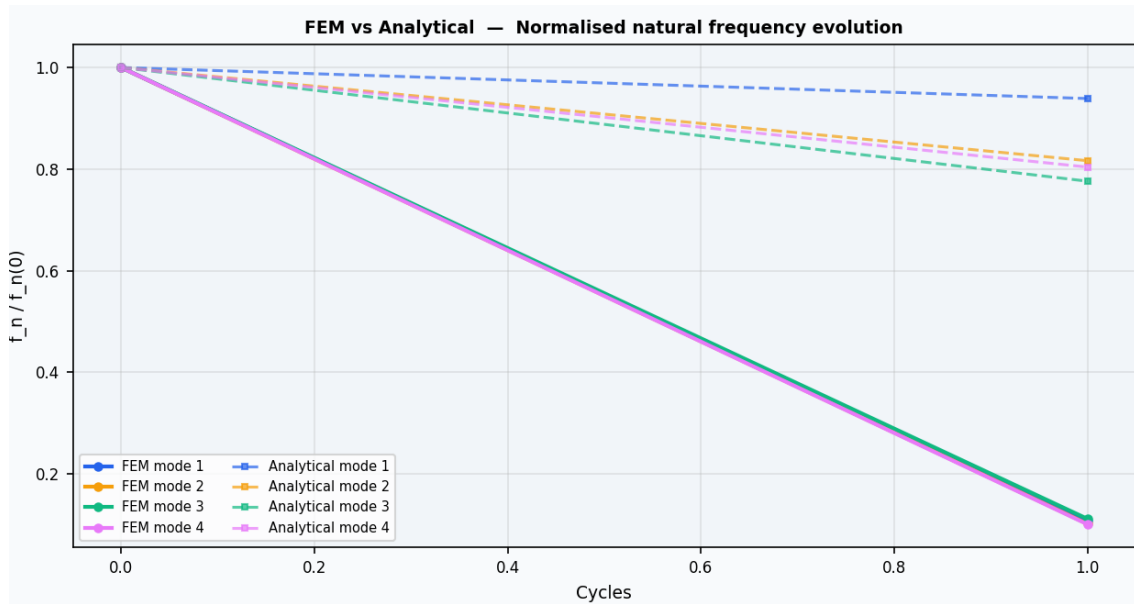


Figure 3.2: Normalised natural frequency evolution $f_n / f_{n,0}$ for the first four modes over the simulated life, comparing Tier 2 (analytical, dashed) against Tier 3 (FEM, solid). As with Figure 3.5, this comparison predates the final specimen definition and uses the design-phase candidate geometry.

3.1.3. Sensor Models and Damage Indicators

The toolkit incorporates physics-accurate transfer-function models for the three sensor types relevant to the original and revised project scope: a contact piezoelectric accelerometer (PCB 352C33), the scanning LDV used in the experimental campaign (Polytec PSV-500), and the quantum photonic vibrometer (QCi QPV) that formed the basis of the original thesis topic. Each sensor model converts the true structural velocity spectrum into a simulated measured signal by applying the sensor's frequency-dependent transfer function and adding calibrated hardware noise, including shot-noise-limited optical noise for the LDV and QPV, and mass-loading effects for the contact accelerometer.

Six damage indicators are evaluated in the simulated environment at each damage increment: natural frequency shift, resonance amplitude shift, modal damping change, phase quadrature deviation, FRF shape correlation, and Modal Assurance Criterion (MAC) drop. For each indicator and each sensor, a Monte Carlo probability-of-detection (POD) analysis estimates the damage level at which detection becomes reliable. These POD curves, generated during the design phase, were used to rank the expected early sensitivity of each indicator and thus informed the priority ordering of the feature vector in the experimental pipeline. Figure 3.3 reports the LDV result, and Figure 3.4 places it alongside the accelerometer and QPV sensor models for a preliminary candidate geometry.

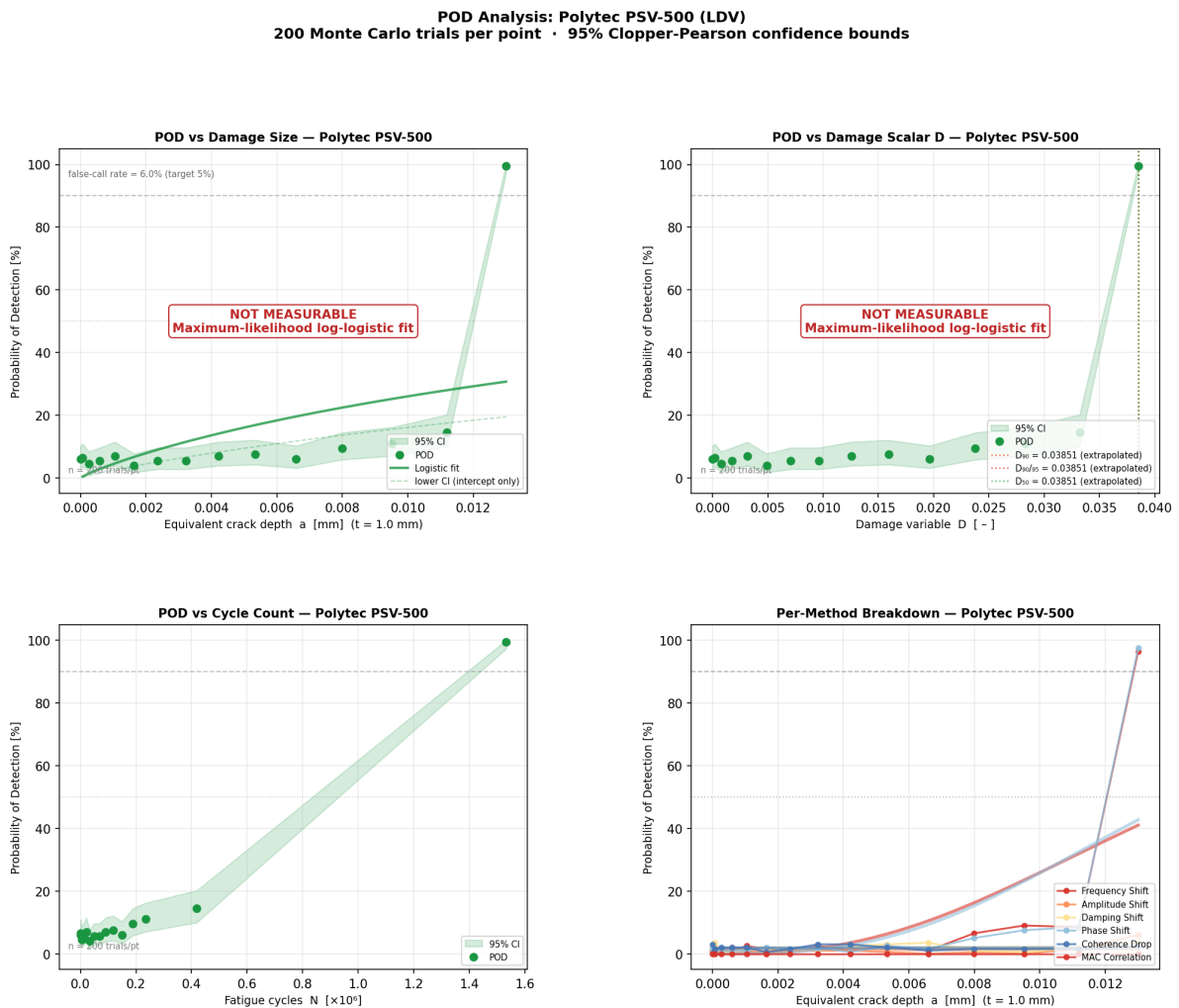


Figure 3.3: Probability-of-detection analysis for the Polytec PSV-500 (LDV) sensor model, from 200 Monte Carlo trials per point with 95% Clopper-Pearson confidence bounds, plotted against equivalent crack depth (top left), the damage scalar D (top right), fatigue cycle count (bottom left), and broken down by individual indicator (bottom right).

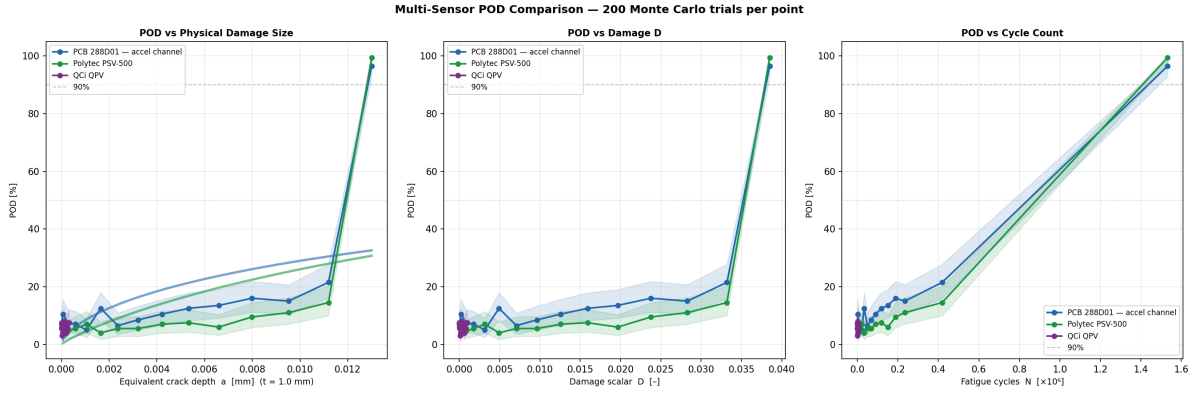


Figure 3.4: Multi-sensor probability-of-detection comparison (200 Monte Carlo trials per point) for the three modelled sensors, against physical damage size (left), damage scalar D (centre), and cycle count (right).

For the purposes of this thesis, only the LDV sensor model outputs are directly compared against experimental data. The QPV model outputs are retained in the toolkit but are not reported as results. They are preserved for the future sensor comparison study identified as a direction for future work in Chapter 7.

3.1.4. Role in the Experimental Campaign

The toolkit's contribution to the experimental campaign can be summarised under three headings.

Specimen and protocol design.

Tier 2 and Tier 3 simulations of the candidate beam geometries predicted the first bending mode frequency as a function of length, width, and thickness, and established that the $196 \times 20 \times 1$ mm geometry would produce a natural frequency of approximately 30 Hz, low enough to avoid excessive shaker force requirements while remaining well above the frequency range dominated by building vibration. Tier 3 simulations, the only tier that resolves a notch as an explicit geometric feature (Section 3.1.2), predicted that a notch-localised stress concentration near the clamped root would produce a detectable MAC drop and curvature anomaly before the frequency shift alone would cross the measurement noise floor; the corresponding predicted mode shape, stress, and damage fields are shown in Figure 3.5. This prediction supplied the initial target rather than a final answer. The specific notch geometry and position that ultimately satisfied this target were arrived at empirically through the Group B load configuration study (Section 3.4.3) rather than fixed by simulation.

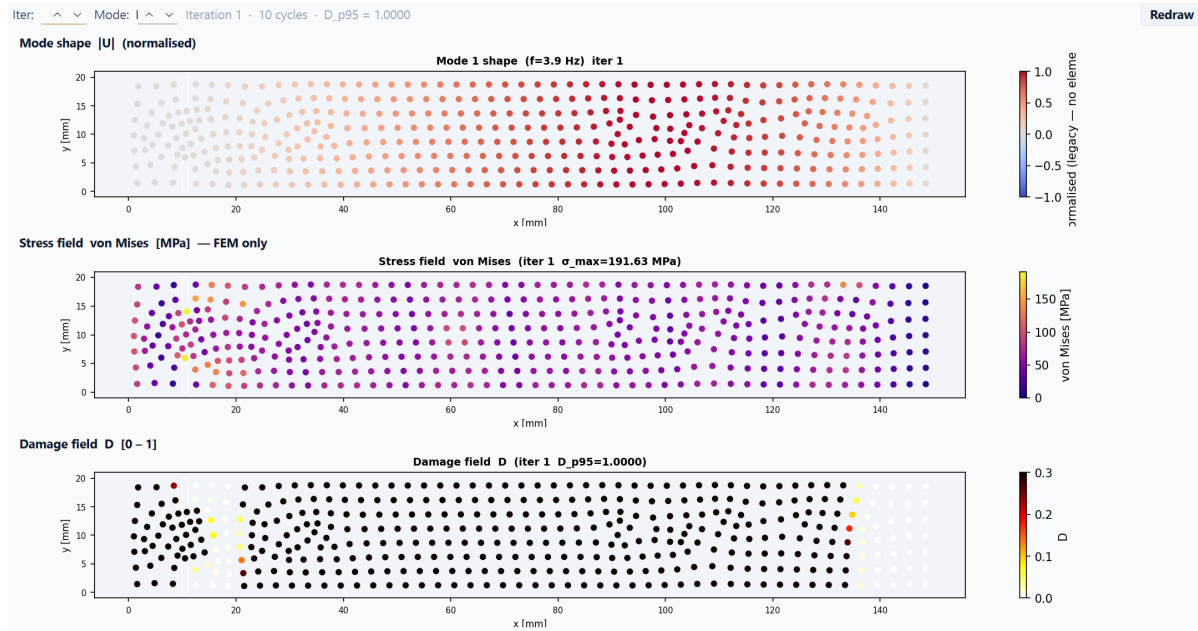


Figure 3.5: Tier 3 (FEM) spatial fields at an early damage increment, computed for the design-phase candidate geometry rather than the specimen finally adopted, since the notch profile was not settled until the Group B study (Section 3.4.3): mode 1 shape (top), von Mises stress field (middle), and CDM damage field D (bottom). All three fields are concentrated at the notch, close to the clamped end (left of each panel), rather than at the beam midspan, consistent with the notch placement described in Section 3.2.2.

Sensitivity analysis.

An OAT and Morris parameter sensitivity analysis was also run during the design phase, intended to rank which material and geometry parameters most strongly affect the final accumulated damage. In practice, this analysis proved uninformative for its intended purpose: on both Tier 2 (Morris) and Tier 3 (OAT), sensitivity was completely dominated by whichever parameter controlled proximity to resonance (excitation amplitude on Tier 2, excitation frequency on Tier 3), with every material parameter showing a normalised sensitivity indistinguishable from zero. Figure 3.6 shows why: damage accumulation is gated by a resonance threshold within the fixed simulation budget, so a parameter combination that keeps the specimen at or near resonance reaches full damage regardless of the specific material properties, while one that detunes the specimen away from resonance accumulates negligible damage regardless of those same properties. This is a documented behaviour of the toolkit's local sensitivity framework when the baseline configuration is not robustly parked at resonance, rather than a genuine finding that material properties do not matter; a re-run with the baseline re-anchored closer to resonance before perturbing material parameters would be needed to obtain a usable ranking. The sensitivity analysis, therefore, did not end up informing the specimen or protocol design decisions, unlike the POD analysis above; those decisions rest instead on the more direct DAF- and stress-threshold arguments described earlier in this subsection.

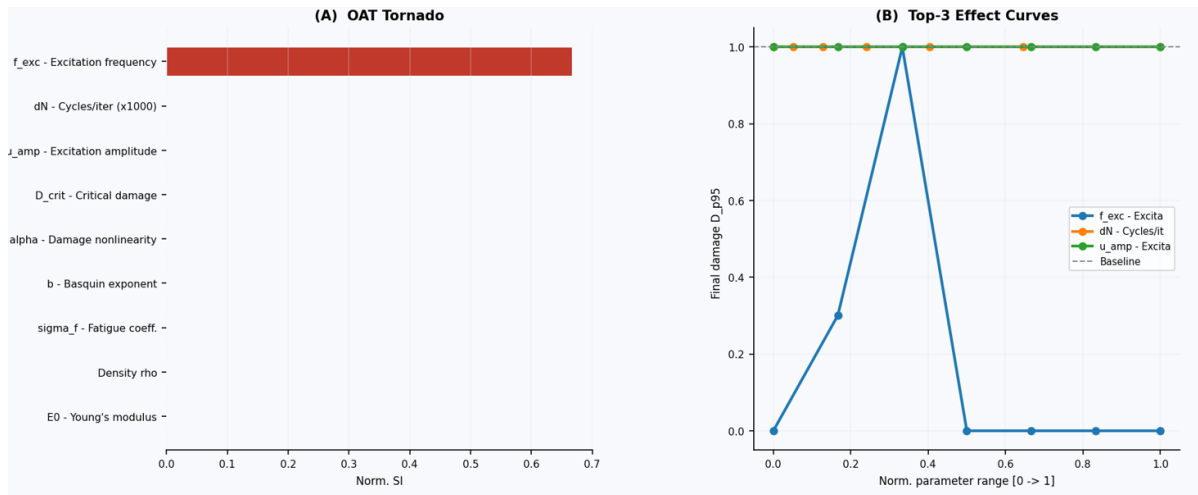


Figure 3.6: Tier 3 OAT sensitivity analysis for final damage D_{p95} . Left: tornado chart, showing excitation frequency as the only parameter with non-negligible normalised sensitivity. Right: the underlying effect curve for the top three parameters, showing the sharp resonance window responsible for this result – damage saturates to its maximum value only within a narrow band of excitation frequency, collapsing to zero outside it.

Post-hoc result interpretation.

After the fatigue campaign was complete, the simulated damage index trajectories were compared qualitatively against the experimental DI curves from Layer 2 of the analysis pipeline (Section 4.3.4), in the same exploratory spirit described above. The toolkit's idealised boundary conditions and, on the analytical tiers, smoothed notch representation mean this is a qualitative check on the CDM-Basquin model's adequacy, not a validation exercise. Where the experimental curves follow the predicted S-shaped logistic trend, this supports the CDM-Basquin physical model as an adequate description of the damage kinematics.

3.2. Specimen Design and Material

The specimen geometry and material were selected to satisfy four interdependent requirements: a first bending mode frequency low enough to be driven by the available microshaker without excessive force, a natural frequency high enough to accumulate fatigue cycles in a practically reasonable test duration, a geometry that produces a well-defined and reproducible crack initiation site accessible to the scanning LDV, and material properties that are well characterised in the open literature to permit comparison with the CDM-Basquin model of the Synthetic Data Toolkit. The final geometry emerged from an iterative process combining Tier 2/Tier 3 toolkit simulations (Section 3.1) with two rounds of preliminary specimen testing (nine specimens in total, Section 3.4): an initial exploration of gross geometry, notch type, and boundary conditions (Group A), followed by a load configuration study that converged on the final notch geometry, position, and drive configuration (Group B).

3.2.1. Material

All ten specimens of the final fatigue campaign were machined from Aluminium 6061-T6 flat bar stock. The primary reason for this choice was practical: material availability, sourced from the faculty's DEMO lab stock. The alloy nonetheless suits the study well for several further reasons. First, its fatigue behaviour in the high-cycle regime is well characterised: the relevant material parameters are available with sufficient accuracy to anchor the toolkit predictions without fitting. Second, Al 6061-T6 is widely used in resonance fatigue studies on metallic beams (Česník et al., 2012; Magi et al., 2016b), providing a direct basis for contextualising the results within the existing literature. Third, its relatively low stiffness ($E_0 = 69$ GPa) and moderate density ($\rho = 2700$ kg/m³) produce first bending mode frequencies in the tens-of-hertz range for millimetre-scale beam geometries, compatible with the shaker force capacity and load cell resolution of the available test rig. The key material properties used in both the toolkit simulations and the experimental data interpretation are summarised in Table 3.1.

Table 3.1: Material properties of Al 6061-T6 used in the Synthetic Data Toolkit and for experimental data interpretation (MatWeb, LLC, 2026).

Property	Symbol	Value	Unit
Young's modulus	E_0	69	GPa
Density	ρ	2700	kg/m ³
Poisson's ratio	ν	0.33	—
Endurance limit*	σ_{end}	97	MPa
Basquin exponent	b	-0.106	—
Stiffness degradation	α	1.5	—

*Aluminium alloys, unlike ferrous metals, do not exhibit a true infinite-life endurance limit; the S-N curve continues to descend gradually well beyond 10^6 cycles. The value reported here follows common engineering practice in treating the fatigue strength at a conventionally defined reference life as a design endurance limit, consistent with published fatigue design data for this alloy (MatWeb, LLC, 2026).

3.2.2. Geometry and Notch Design

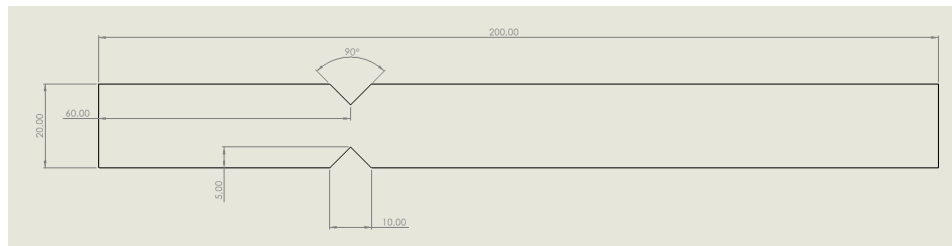
The final specimen geometry, dimensioned in Figure 3.7a and shown as manufactured in Figure 3.7b, is a thin rectangular beam of total stock dimensions $196 \times 20 \times 1$ mm (length $\times$ width $\times$ thickness). The specimen is clamped over the first 45 mm from one end, leaving an effective cantilevered length of approximately 151 mm. This geometry places the first bending mode at approximately 30 Hz under cantilever boundary conditions, and, for each Group C specimen individually, by the baseline modal identification performed at the start of its own Phase 3 run (following the S1 methodology, Section 3.4.2). At 30 Hz, one fatigue block of $N_{\text{block}} = 7,500$ cycles corresponds to approximately 4 minutes 10 seconds of harmonic loading (Section 3.5.1), giving a total Phase 4 campaign duration per specimen on the order of one to a few hours depending on the specimen's fatigue life.

The low thickness of 1 mm deserves specific mention. It implies that geometric nonlinearity, namely, large-deflection effects in which the stiffness matrix becomes amplitude-dependent, can appear at surprisingly low vibration amplitudes for a specimen of this slenderness. This constraint was identified through the nonlinearity validation test (S4) described in Section 3.4, which found the onset of measurable nonlinearity at a drive level of approximately 0.3 V for a load placement 6 cm after the clamped edge; its implications for the Phase 3 and Phase 4 excitation levels are discussed in Section 3.5.3 and in more detail in Chapter 6.

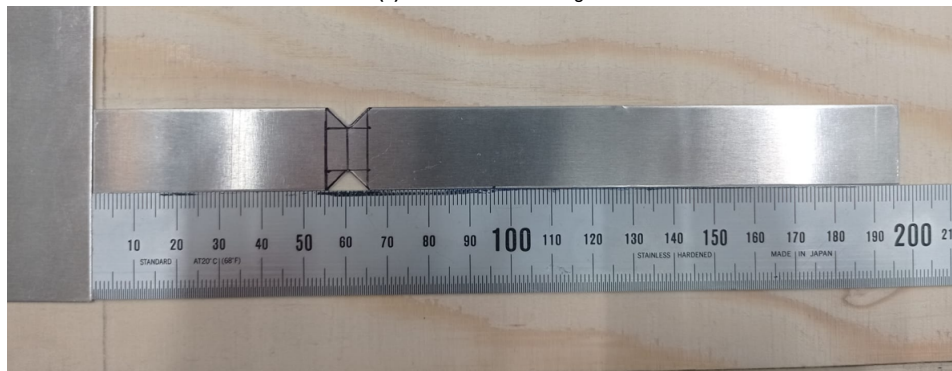
The beam carries symmetric side notches serving two purposes: fixing the crack initiation site at a known location, and placing the resulting damage where the tracked indicators are most likely to respond to it. Both requirements point to the same region of the beam. Bending moment in the first cantilever mode is greatest at the clamped root and decays toward the free end, so a notch placed as close to the clamp as practical sees the highest cyclic stress and initiates soonest. That same region carries the highest mode-shape curvature, which is the quantity the spatial indicators respond to, so damage localised there perturbs the measured mode shape more strongly than damage further along the beam. The notches were accordingly positioned 60 mm from the clamped end, 15 mm beyond the clamp face: as close to the root as the clamp geometry allowed, while leaving clearance for the LDV scan grid to be laid out around the notch along the beam neutral axis (Section 3.3.2).

Each notch is an isosceles right-triangular cutout, its 10 mm base on the beam edge running along the beam axis and its apex 5 mm into the 20 mm width (Figure 3.7). Cutting both edges leaves a net ligament of 10 mm at the notch plane, halving the load-bearing width relative to the gross section, and the right-angled apex gives an included notch angle of 90° . The notches were produced by cutting stamping rather than milling, which forms sharp 90° edges at the notch root. No root radius was therefore specified, and none was measured on the finished specimens, so the elastic stress concentration factor K_t is not known analytically for this geometry. The only estimate available is the one implicit in the Tier 3 finite element model of the Synthetic Data Toolkit (Section 3.1.2), which resolves the notch as an explicit meshed feature and returns the local von Mises field from which the concentration follows; that estimate inherits the idealised sharp-cornered notch of the CAD geometry and is used here to reason about relative stress distribution, not as a calibrated K_t for the manufactured part. The notch is accordingly treated throughout as an empirically calibrated initiation feature. The V-profile replaced

an initially planned rectangular slot, a Group A design that carried into the early Group B specimens before proving unreliable to manufacture with consistent geometry; the replacement was settled during the Group B load configuration study (Section 3.4.3). In service this notch channels the fatigue crack across the beam cross-section at a known, fixed location, which is what allows the LDV scan grid to be positioned relative to it in advance (Section 3.3.2).



(a) Dimensioned drawing.



(b) Photograph of a machined specimen.

Figure 3.7: Final specimen geometry (Al 6061-T6, clamped over the first 45 mm): (a) dimensioned drawing; (b) a manufactured specimen. The drawing dimensions the 200 mm beam originally specified; the stock plate ultimately obtained yielded a 196 mm beam, and all reported results refer to that dimension.

“latex

3.2.3. Specimen Population

The experimental programme comprised three main groups of specimens, each associated with a distinct stage in the development of the final test configuration. The third group was subsequently divided into two subgroups according to its role in the final analysis. Two additional groups, produced after completion and analysis of the main campaign, were used for diagnostic follow-up and are described separately below.

- *Specimen-design exploration (Group A, $n = 4$).* Four specimens spanning a range of gross geometries, notch types, and boundary conditions were used to establish whether a specimen could be fractured reliably with the available shaker and fixture, and to identify a configuration that would fracture consistently at a targeted location. The exploration included composite specimens, which were ultimately abandoned in favour of the metallic specimens used in the remainder of the study because they proved more difficult to fracture within the available actuator capacity. Further exploration was not justified given the project's overall time constraints (Section 1.3).

Group A was independent of the sensor-characterisation tests described in Section 3.4. Tests T1–T4 were conducted without a specimen mounted, with the LDV measuring the load cell/shaker armature directly, while S1–S5 required the final specimen geometry, which had not yet been established at this stage. Group A therefore served exclusively as a specimen-design study and did not contribute to either test battery. The range of geometries explored is shown in Figure 3.8a.

- *Load configuration specimens (Group B, $n = 5$).* Five specimens fabricated to the geometry emerging from the Group A exploration were used for the iterative load-configuration study re-

ported in Section 3.4.3. This study established the rig and clamping configuration, sample placement, excitation level, and final notch geometry.

Group B specimens also carried the S1–S5 sample-characterisation battery (Section 3.4.2) and the T4d surface-finish sensor test (Section 3.4.1), the latter being the only T1–T4 sub-test requiring an actual specimen surface. Figure 3.8b shows the resulting specimens after testing, alongside the Group A specimens shown in Figure 3.8.

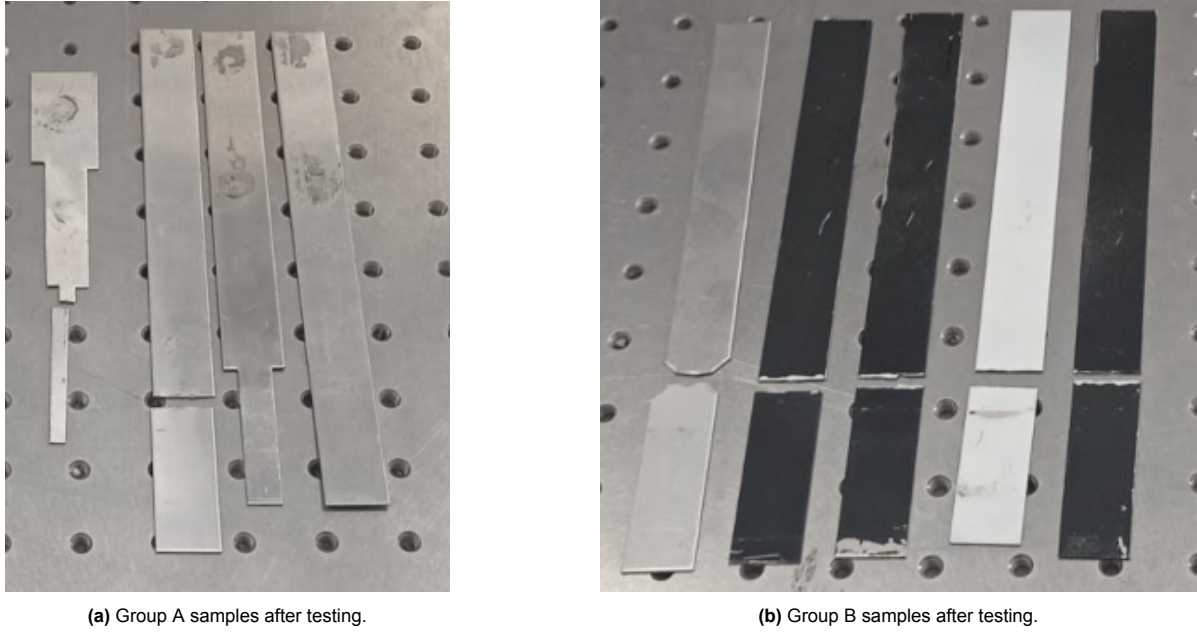


Figure 3.8: Specimens used in the pre-test campaign to explore specimen designs, loading, and the broader experimental setup.

- *Final campaign specimens (Group C, $n = 10$).* Ten specimens, all machined from the same Al 6061-T6 flat bar stock to nominally identical dimensions, constitute the primary experimental dataset. All ten underwent identical Phase 3 (healthy baseline) testing and Phase 4 (fatigue loading) excitation under the fixed-frequency protocol. The population was divided into two subgroups according to their role in the validation strategy.

Group C1 ($n = 8$) – complete run-to-failure specimens. Eight specimens were tested under the standard Phase 4 protocol without interruption until physical fracture occurred. They constitute the population-level dataset: the complete $D(\tau)$ trajectories from Layer 2, the indicator sensitivity rankings of Layer 1, and the Gaussian Process training set of Layer 3 are all based exclusively on this subgroup. The total cycles to fracture, N_f , varied between specimens because of natural inter-specimen scatter in surface condition, crack initiation timing, and minor clamping variability. This scatter is characterised as part of the Layer 2 damage-index analysis and is normalised through the cycle-fraction axis $\tau = N/N_f$ before cross-specimen comparison.

Group C2 ($n = 2$) – interrupted validation specimens. Two specimens followed a deliberately different Phase 4 protocol. Loading was interrupted at the first damage alarm raised by the live-monitoring system (Section 4.6) rather than being continued to fracture. Each specimen was characterised by optical microscopy before Phase 4, re-inspected immediately after the interruption, and subsequently loaded to final fracture. This enabled its fracture surface to be compared with those of the Group C1 specimens (Section 3.5.5).

Because the interruption point was determined by the live-detection logic rather than by a fixed protocol parameter, Group C2 was not pooled with Group C1 for the Layer 1 sensitivity analysis, Layer 2 inter-specimen comparison, or Layer 3 Gaussian Process training. Instead, it provided an independent, microscopy-based validation track for the detection logic, described in Section 3.5.5.

Diagnostic specimens (Groups D1 and D2).

Groups D1 and D2 were produced after the main campaign had been completed and analysed. They were introduced to investigate two causes of measurement scatter identified in Chapter 6: insufficient fixity in the clamping arrangement and an impure load-transfer path between the shaker and specimen. The D groups therefore constitute a diagnostic follow-up rather than an extension of the population-level dataset, and their results are not pooled with Group C.

Both D groups used a shortened Al 6061-T6 beam measuring $130 \times 20 \times 1$ mm. The notch geometry was unchanged from Group C, with its centre located 55 mm from the clamped end, or 15 mm beyond the clamp face. The specimens were clamped over the first 40 mm, leaving a 90 mm cantilevered length compared with 151 mm for Group C. This increased the first bending-mode frequency from approximately 30 Hz to approximately 100 Hz, allowing a given number of cycles to be accumulated in a shorter test.

The shortened geometry also means that the D-group configuration changes more than the clamping and load-transfer arrangements alone. The increased beam stiffness reduces the relative contribution of clamp compliance, while the shorter test duration reduces the time available for fixture drift to accumulate. The resulting evidence is therefore corroborative rather than isolating. This limitation is considered explicitly in the interpretation of the results in Section 6.1.1.

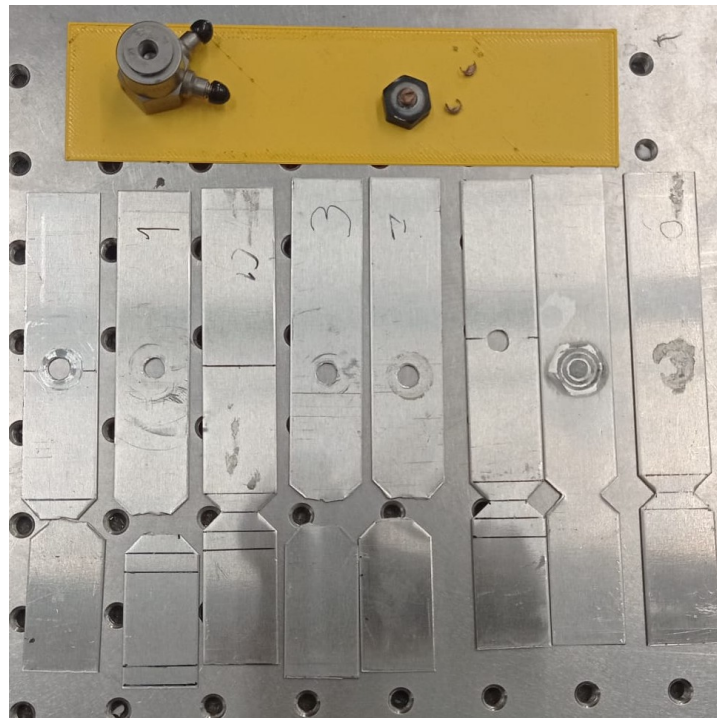


Figure 3.9: Group D specimens after testing. At $130 \times 20 \times 1$ mm, the beam is substantially shorter than the Group C geometry of Figure 3.7; the notch geometry is unchanged, and the through-hole near the upper end carries the bolted load-transfer connection selected in the Group D1 study. Every specimen shown has separated at the notch.

- *Boundary condition and load transfer study (Group D1, $n = 3$).* Three specimens were tested iteratively, following the approach used for Group B rather than a fixed test matrix. The study addressed clamping and load-transfer arrangements in sequence. The clamping trials used a series of printed inserts to control the height and depth at which the specimen was seated within the clamp jaws. The retained arrangement seated the specimen directly against the flattest lower face of the clamp, with clamping torque held constant across runs and specimens to ensure reproducible seating.

The load-transfer trials compared adhesive tape at two coverage levels, printed fixtures, structural adhesive, and a drilled-hole bolted connection. The bolted connection was retained because it

maintained contact throughout the test without debonding and introduced a fixed, known mass rather than the variable mass associated with tape and adhesive. Group D1 was analysed using the LDV Analysis Suite alone, following the S1–S5 methodology of Section 3.4.2; no fatigue-campaign analysis layer was applied.

- *Repeat fatigue tests (Group D2, $n = 5$)*. Five specimens were subjected to the Group C1 fatigue protocol of Section 3.5, using the configuration selected through Group D1. All were driven at 1 V and tested to fracture. The block structure was shortened relative to Group C1: the 10 s broadband scan was retained, while the harmonic-loading stage was reduced so that a complete block occupied approximately one minute rather than the 5 min 40 s specified in Section 3.5.1. The higher drive frequency allowed a given cycle count to be accumulated more rapidly, while the available testing time imposed the shorter block duration.

Block counts to fracture varied across the five specimens. Two survived 7 and 5 blocks, respectively, while two others survived only 2 and 1 blocks. The fifth specimen was tested across two separate days and is excluded from the reported set because the remounting event could contaminate a comparison intended to assess the mounting arrangement itself; this specimen fractured after 6 blocks. Consequently, four specimens enter the results reported in Section 5.8. Only Layer 0 of the analysis pipeline was applied to Group D2; Layers 0.5–3 were not.

The results from Groups D1 and D2 are reported in Section 5.8. Their interpretation is restricted to R1 and the data-quality assessment in Section 6.1.1; they do not contribute to the indicator sensitivity ranking (R2), damage-index and prognostic-horizon analysis (R3), or S–N characterisation (R4), which remain based on Group C1.

The specimen groups and their roles are summarised in Table 3.2.

Table 3.2: Summary of specimen groups, their purpose, and their role in the experimental campaign.

Group	n	Geometry / material	Primary purpose
A – Specimen-design exploration	4	Various geometries, notch types, and boundary conditions	Establish a geometry/notch combination that fractures reliably at a targeted location
B – Load configuration	5	Al 6061-T6, converging on final $196 \times 20 \times 1$ mm geometry with V-notches	Iterative rig, placement, and excitation-level tuning; final notch geometry settled here; also carries S1–S5 and T4d
C1 – Final campaign (complete)	8	Al 6061-T6, $196 \times 20 \times 1$ mm, V-notches	Primary dataset: Phase 3 baseline and Phase 4 fatigue, run to fracture; population-level Layer 1–3 analysis
C2 – Interrupted validation	2	Al 6061-T6, $196 \times 20 \times 1$ mm, V-notches	Phase 4 interrupted at first live-monitoring alarm; microscopy-based validation of the detection logic
D1 – Boundary/load-transfer study	3	Al 6061-T6, $130 \times 20 \times 1$ mm, unchanged notch geometry	Iterative evaluation of clamping and load-transfer arrangements; S1–S5 analysis only
D2 – Repeat fatigue tests	5	Al 6061-T6, $130 \times 20 \times 1$ mm, unchanged notch geometry	Repeat fatigue tests under the revised configuration; Layer 0 analysis only

3.3. Experimental Setup and Instrumentation

The general configuration – a cantilevered beam specimen, electrodynamic shaker excitation, an instrumented load cell as the force reference, and a non-contact optical response sensor – follows the broad approach established in the resonance fatigue testing literature: base-excited beam and plate specimens driven near resonance by an electrodynamic shaker (George et al., 2004), and the pairing of a shaker/load-cell drive with scanning LDV response measurement specifically for phase- and mode-shape-based damage monitoring (Česnik et al., 2012; Di Maio et al., 2018; Magi et al., 2016b). The rig built for this thesis adapts that general approach to the constraints of the available laboratory equipment (Section 3.3.1) rather than reproducing any single prior setup directly.

The experimental rig was designed around three non-negotiable constraints imposed by the monitoring objectives of the study: the response sensor must be non-contact to avoid mass-loading the specimen's modal parameters throughout the fatigue life; the force input and its associated reference signal must be measured simultaneously with the structural response at every block, to enable the phase-difference indicator and coherent FRF estimation; and the measurement system must maintain stable calibration over repeated fatigue blocks spanning hours of continuous operation. These constraints, together with the scan requirement for spatial mode-shape indicators, led to the configuration described below. Figure 3.10 provides a schematic overview, while Figure 3.11 shows the updated arrangement after Group D testing.

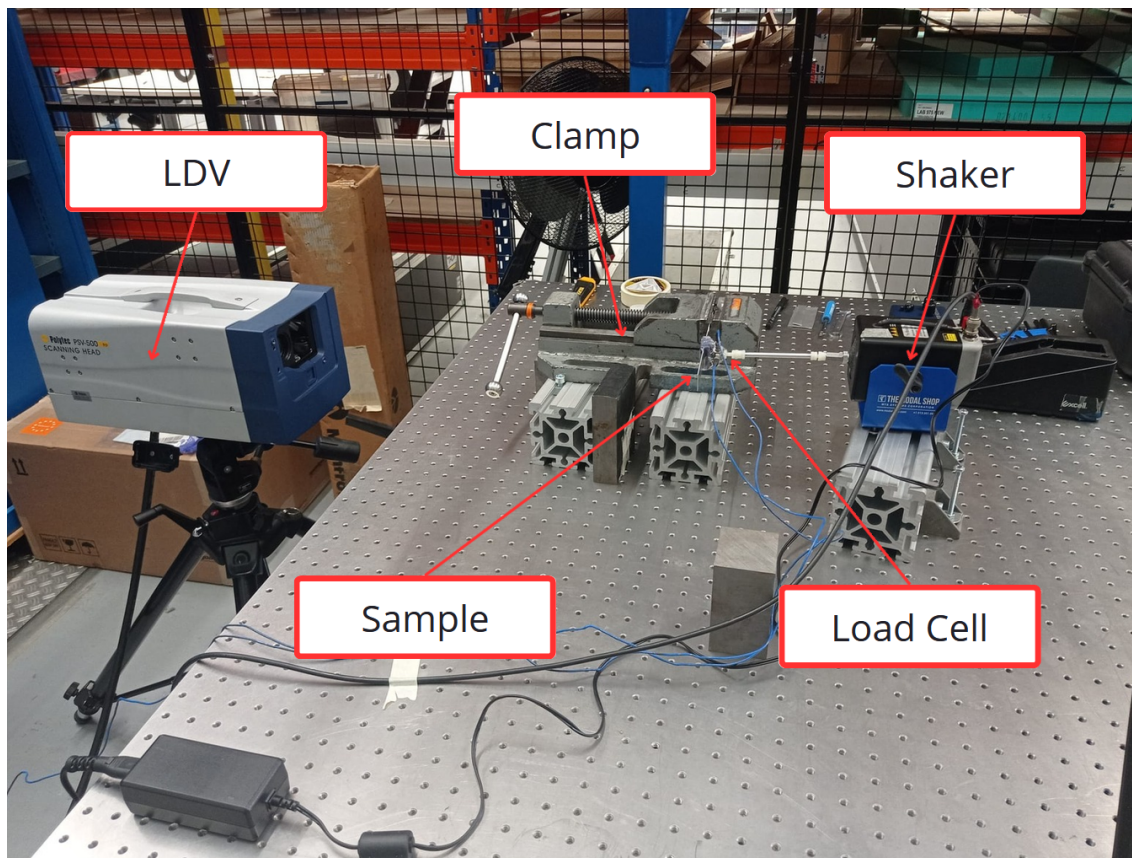


Figure 3.10: Schematic and photograph of the experimental setup. The specimen is clamped horizontally at one end; the microshaker applies transverse excitation near the notch, toward the free end, via an instrumented load cell. The scanning LDV measures the out-of-plane velocity response along the beam neutral axis from a fixed stand-off distance.

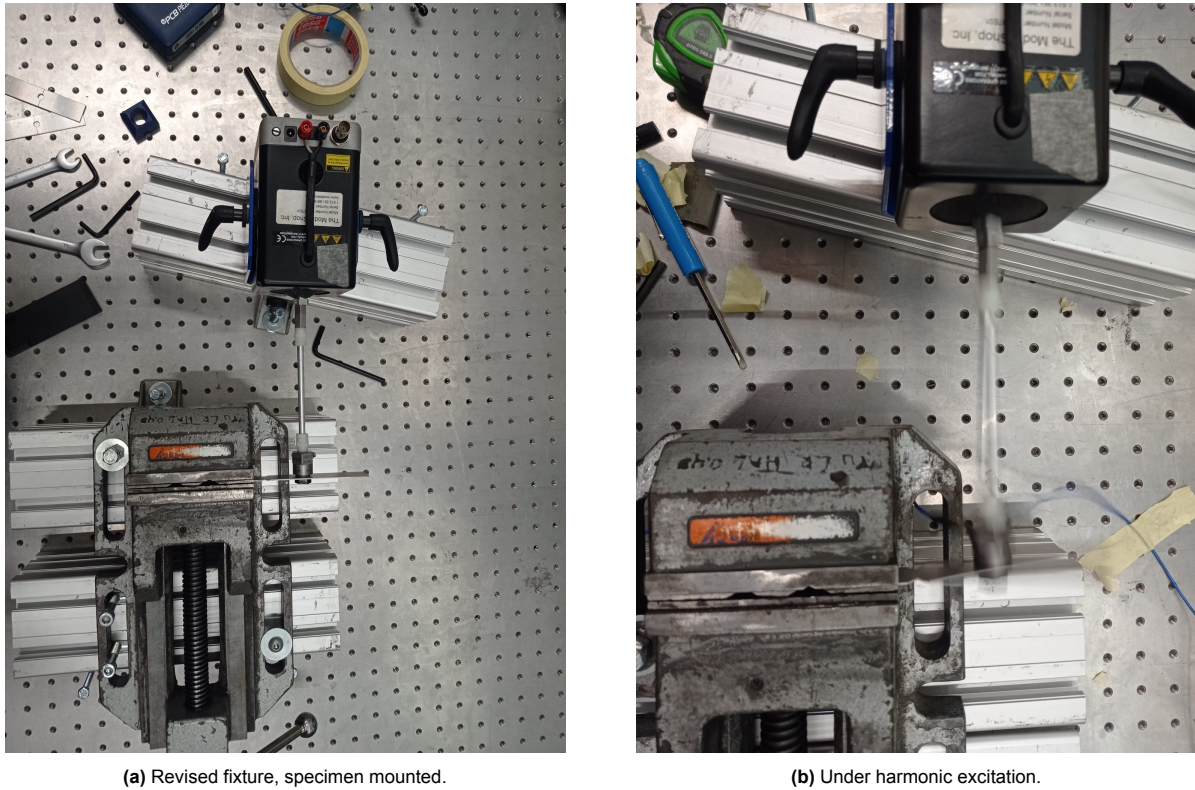


Figure 3.11: Revised boundary condition and load transfer arrangement arrived at in the Group D1 study (Section 3.2.3) and used for Group D2, shown for comparison against the main-campaign configuration above. In (b) the specimen is blurred by its own motion during excitation.

3.3.1. Fixture and Loading Configuration

The specimen is mounted horizontally as a cantilever: one end is clamped in a rigid fixture bolted to the laboratory table, and the opposite end is free. Clamping consistency is one of the principal sources of inter-specimen variability in resonance fatigue experiments (Česnik et al., 2012; Di Maio et al., 2018), and its control is a prerequisite for meaningful cross-specimen comparisons of modal parameter evolution. To minimise this variability, the clamping block is tightened as consistently as possible across specimens by controlling the number of rotations applied to the clamp.

Excitation is provided by a PCB Piezotronics K2007E01 miniature electrodynamic shaker, referred to throughout as the microshaker, rated to 31 N (7 lbf) peak sine force with a 13 mm peak-to-peak stroke and a usable frequency range extending to 9 kHz (PCB Piezotronics, Inc., 2026b), comfortably above the approximately 30 Hz operating frequency of this campaign. Most comparable resonance fatigue rigs reported in the literature use larger electrodynamic shaker tables mounted directly beneath or adjacent to the specimen (Česnik et al., 2012; Magi et al., 2016b). A shaker table of that scale was not available for this project, so the microshaker was instead mounted approximately 17 cm from the specimen and connected to it via a rigid extension arm.

Arriving at a stable mounting for the microshaker itself required resolving two further practical problems during the Group B load configuration study (Section 3.4.3). First, the shaker's mounting-hole pattern did not match the optical table's grid, and an intermediate aluminium extrusion strut was introduced to provide a rigid, correctly-spaced mounting interface between the two. Second, an early vertical mounting arrangement made it difficult to keep the drive arm perpendicular to the beam while still allowing its position to be adjusted between configurations; this was resolved by switching to the horizontal mounting configuration shown in Figure 3.10, which keeps the arm perpendicular to the beam by construction while remaining straightforward to reposition along the beam axis.

This overall approach adds complexity relative to a direct shaker-table mount: the arm introduces its own modal properties into the drive path, coupling the shaker's dynamics with those of the beam rather

than applying a clean base excitation, and the arm-to-specimen connection deflects more than a direct shaker-table mount would, both effects increasing the noise floor and the risk of drive-side nonlinearity relative to the more rigid configurations typically reported in the literature. This is treated as a limitation specific to the present rig and is revisited in Section 6.1.1 of Chapter 6, where the indicator noise levels it produces are interpreted.

The shaker armature (via the extension arm) is connected to the upper face of the beam via a PCB Piezotronics 288D01 miniature impedance sensor (S/N 3896), which is bonded to the beam surface using a rigid adhesive, at a distance of approximately 4–5 cm from the notch. This distance was varied deliberately across the earliest Group C1 specimens (4, 4.5, then 5 cm) before being held fixed at 5 cm for the remainder of the campaign: informed by the Group B load configuration study, moving the attachment point changes the effective moment arm and therefore the stress delivered to the notch for a given drive voltage, giving a second, independent lever (alongside the Phase 4 drive voltage itself, Section 3.5.3) for broadening the range of fatigue lives sampled across the Group C1 population ahead of the Basquin S-N fit (Section 4.5). The 288D01 combines a force sensor (100 mV/lbf, ± 50 lbf range) and an accelerometer (100 mV/g, ± 50 g range) in a single package (PCB Piezotronics, Inc., 2026a), which is what allows it to serve a dual function: the force channel is the input channel for FRF estimation, and the acceleration channel provides the shaker-side reference signal used for the phase-difference indicator ϕ_{diff} described in Section 4.1.6. The attachment location, near the notch on the free-end side, was determined during the load configuration study (Section 3.4) to be the configuration that produced consistent, reproducible fracture at the notch within the target number of fatigue blocks, while maintaining adequate coherence across the frequency range used for modal identification.

The load cell mass is acknowledged as a perturbation to the specimen dynamics, but its effect is fixed and constant across all blocks and all specimens, entering only as a static offset in the absolute value of f_n rather than as a factor affecting its evolution. (The load cell's own acceleration-reference cable did become disconnected partway through the Specimen 7 test; this data-quality note is recorded alongside that specimen's results in Chapter 5.)

3.3.2. Scanning Laser Doppler Vibrometer

Structural velocity measurements are acquired with a Polytec PSV-500 scanning Laser Doppler Vibrometer equipped with a DV-04 velocity decoder. The instrument operates on the principle of heterodyne interferometry: a helium-neon laser beam is split into a measurement arm directed at the specimen surface and a reference arm frequency-shifted by a Bragg cell. The Doppler-shifted return beam and the reference beam are recombined at a photodetector, and the demodulated photocurrent is decoded to yield the instantaneous surface velocity $v(t)$. The instrument's rated velocity range extends from approximately $0.01 \mu\text{m s}^{-1}$ up to 30 m s^{-1} , over a bandwidth of DC to several megahertz depending on decoder configuration (Polytec GmbH, 2026), comfortably spanning the velocity amplitudes produced by the specimens under both Phase 3 and Phase 4 loading conditions.

The LDV is mounted on a rigid tripod at a fixed stand-off distance of approximately 660 mm from the beam surface, aligned perpendicular to the beam axis to maximise sensitivity to out-of-plane bending velocity. The motorised scan mirror deflects the laser beam sequentially to each programmed grid point along the beam neutral axis, acquiring the FRF at each position autonomously before stepping to the next.

The scan grid consists of 9 points spanning a 20 mm region, beginning 10 mm beyond the notch on the free-end side and extending away from it, giving a point spacing of 2.5 mm. The notch itself is deliberately excluded from the measured region. Two considerations motivate this offset. The notch is a machined discontinuity, and the abrupt change in section it introduces would dominate any spatially differentiated quantity computed across it, so curvature estimated over a grid spanning the notch would report the geometry rather than the damage. In addition, once a crack opens at the notch root the local surface becomes optically unfavourable, with the edges of the opening scattering the incident beam and degrading the return signal at precisely the points a notch-spanning grid would rely on most. Measuring immediately upstream of the notch instead captures the mode-shape perturbation that a growing crack induces in the surrounding material, while keeping every scan point on undamaged, optically stable surface for the whole test.

The consequence is that the spatial indicators detect damage through its effect on the beam's deflected shape near the notch rather than by imaging the crack itself, and any localisation they provide is relative to that measured window rather than absolute. This nine-point grid also falls below the 20–30 points identified in the pipeline documentation as the recommended density for reliable modal curvature estimation via central finite differences on beams of this length. The reduced grid was adopted because of practical file-size and scan-time constraints on the multi-hour, multi-specimen campaign (Section 3.3.3), the same constraint that also drove the sample-rate reduction described there. The consequence – coarser curvature and mode-shape-based spatial indicators (MAC, CDI, MSE-TVD, spatial phase scatter) than the pipeline was originally designed to support – is acknowledged here as a limitation of the present campaign and is revisited in Section 6.1.5 of Chapter 6 when the spatial-indicator results are interpreted. The scan grid coordinates are stored in the HDF5 data file alongside the measured FRF data, enabling reproducible alignment between measurement epochs.

3.3.3. Signal Conditioning and Data Acquisition

All analogue signals (LDV velocity output and load cell force channel) are conditioned and digitised by the PSV-500 front-end electronics. There is no separate external DAQ unit. Time-series data are extracted from the PSV-500 system and post-processed offline.

The data acquisition system samples both channels simultaneously at $f_s = 10,000$ Hz. This rate was reduced from an initially planned 100 kHz during the design phase, once file-size growth over a multi-hour, multi-specimen campaign made the higher rate impractical. At 10 kHz, the Nyquist frequency of 5 kHz remains comfortably above the frequency range of interest (the first three bending modes, expected below 1500 Hz for this geometry), so no additional decimation is applied, and all spectral analysis in this thesis operates directly on the acquired 10 kHz signal. Raw time-series data are written to tab-delimited text files for each measurement phase within each block: one file per sensor per harmonic excitation stage, and one file per scan point per sensor for the broadband scanning stage. This file structure is described in detail alongside the signal processing pipeline in Section 4.1.

The instrumentation used in this study is summarised in Table 3.3.

Table 3.3: Summary of instrumentation used in the experimental campaign.

Component	Make / model	Role
Scanning LDV	Polytec PSV-500 (DV-04 decoder)	Non-contact velocity measurement, 0.01 $\mu\text{m/s}$ –30 m/s range (Polytec GmbH, 2026); broadband scan for mode shapes; integrated data acquisition
Microshaker	PCB Piezotronics K2007E01	Harmonic and broadband excitation via an extension arm (Section 3.3.1); 31 N peak sine force, 9 kHz bandwidth (PCB Piezotronics, Inc., 2026b)
Impedance sensor	PCB Piezotronics 288D01 (S/N 3896)	Combined force input channel (± 50 lbf) and shaker-side acceleration reference ($\pm 50g$) for ϕ_{diff} (PCB Piezotronics, Inc., 2026a)
Data acquisition	Integrated in the Polytec PSV-500 front end	Simultaneous sampling at 10,000 Hz; file output
Clamping fixture	Standard clamping block, lab table-mounted	Cantilever boundary condition

3.4. Pre-Campaign Characterisation

Before the main fatigue campaign could begin, three stages of preliminary work were completed: sensor characterisation (T1–T4, conducted without a mounted specimen except for T4d), specimen and sample characterisation (S1–S5, conducted on Group B specimens), and an iterative load configuration

study, also on Group B specimens, that converged on the excitation amplitude, attachment position, and clamping arrangement capable of producing consistent and reproducible fatigue fracture within the target number of blocks. All three stages were conducted using the LDV Analysis Suite, a purpose-built desktop application implementing the test protocols described below. The outputs of these stages directly informed the Phase 3 and Phase 4 protocol parameters and are treated here as prerequisite characterisation work rather than primary experimental results.

3.4.1. Sensor Characterisation (Tests T1-T4)

Four sensor tests characterised the intrinsic measurement performance of the LDV and its drive chain before any specimen-level testing began. With the sole exception of T4d, none of these tests used a mounted beam specimen at all: the LDV was pointed directly at the load cell/shaker armature, isolating the sensor and drive-chain performance from any structural response. T4d, which specifically tests sensitivity to surface material and finish, necessarily required an actual specimen surface and was carried out later, on Group B specimens (Section 3.2.3), rather than at this stage.

- *T1 – Noise Characterisation.* The LDV velocity noise floor was measured under quiescent conditions (shaker off, no mechanical excitation) by computing the power spectral density of the velocity output over a sustained acquisition. The resulting noise floor spectrum establishes the minimum detectable velocity amplitude at any frequency and provides the reference below which no damage-induced indicator drift can be trusted. This test was run first, before any excitation, to ensure that the baseline represents purely instrumental noise uncontaminated by residual structural vibration. The measured noise PSD and cumulative RMS noise for both the LDV and the reference accelerometer are reported in Appendix A (Figure A.1); the LDV noise floor sits several orders of magnitude below the accelerometer's across the full bandwidth.
- *T2 – Minimum Detectable Amplitude.* Rather than a continuous amplitude sweep, T2 characterises detection performance at three fixed frequency points (0.1, 10, and 1000 Hz), at the lowest amplitude practically achievable from the shaker, via a frequency sweep across these three points. This brackets the 30 Hz Phase 3/4 drive frequency without testing it directly, bounding the minimum scan excitation amplitude usable during the broadband sweep in Phase 3 and Phase 4. The cross-frequency metric comparison is reported in Appendix A (Figure A.2).
- *T3 – Broadband and Harmonic Excitation Quality.* T3a and T3b validate the two distinct excitation types used within a single Phase 3/4 block: T3a characterises the broadband scanning excitation used for FRF estimation and modal identification (coherence, FRF magnitude variability, noise-to-signal ratio, at the intended scan amplitude), while T3b characterises the harmonic excitation used for fatigue loading (total harmonic distortion and noise floor as a function of input amplitude). Together, they establish the upper amplitude bound, above which the response becomes non-linear, and the lower bound, below which FRF estimation quality degrades, for each of the two excitation regimes actually used in the protocol.
- *T4 – Disturbance Sensitivity.* Four sub-tests characterised how the LDV noise floor and signal-to-noise ratio respond to practical disturbances encountered in the laboratory: optical power variation (T4a), stand-off distance changes (T4b), beam alignment angle (T4c), and, on Group B specimens, surface material or finish (T4d). The T4 results informed the choice of not using a surface paint treatment and confirmed that the stand-off distance and alignment established during rig setup would maintain adequate measurement quality throughout the fatigue blocks, even as minor thermal drifts in the optical path occurred.

The key output of the T1–T4 battery is a sensor performance report consolidating the noise floor, minimum detectable velocity, usable amplitude range, and sensitivity to practical disturbances for the specific LDV installation in the test laboratory. These quantities set the detection threshold for the anomaly detection layer (Section 4.1.7) and define the measurement margins within which the indicator sensitivity rankings produced by Layer 1 are valid. The full per-test reports, including the T3 and T4 result sets not reproduced here, are given in Appendix A.

3.4.2. Specimen and Sample Characterisation (Tests S1--S5)

Five sample tests were conducted on Group B specimens (Section 3.2.3), fabricated to the geometry converging out of the Group A exploration, to establish baseline modal properties and validate that the planned test conditions satisfy the linearity requirements of the modal analysis pipeline. The tests were executed in a fixed order, as each builds on the outputs of the preceding one. The figures shown below are drawn from one of the earliest Group B tests, conducted before the final 30 Hz drive frequency and V-notch geometry were settled; the specific numerical values shown (e.g. a 25 Hz drive frequency) illustrate the methodology and should not be read as the final campaign parameters, which are given in Section 3.4.3.

- **S1 – Baseline Modal Characterisation.** The natural frequencies, damping ratios, and effective modal masses of the specimen were identified under broadband white noise excitation at low amplitude, using the Least Squares Complex Frequency (LSCF) domain estimator applied to the grand-averaged H_1 FRF computed from three repeated acquisition runs (Allemang & Avitabile, 2022; Ewins, 2000). The LSCF stabilisation diagram, illustrated in Figure 3.12, was inspected, and physical poles were confirmed manually before the residue computation. The S1 output provides the modal seeds used by all downstream tests and, critically, the initial resonance frequency $f_{n,0}$ that becomes the fixed drive frequency for Phase 4.

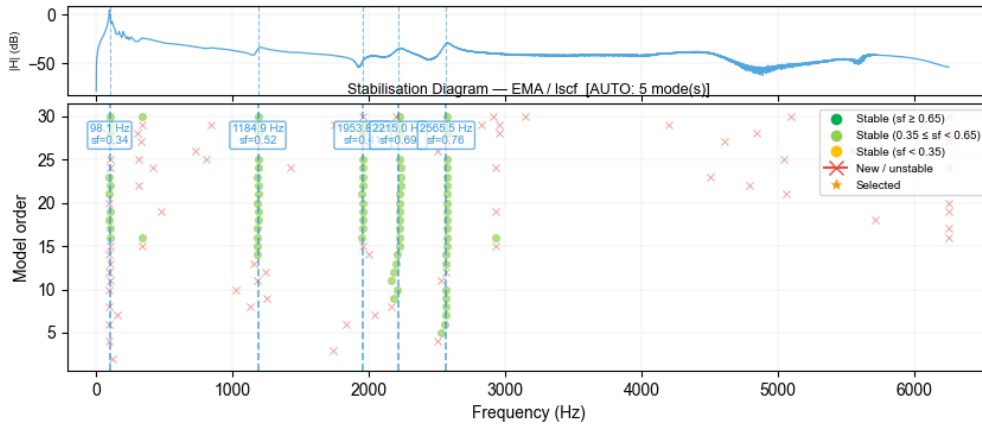


Figure 3.12: S1 baseline modal characterisation: LSCF stabilisation diagram (bottom) with the corresponding FRF magnitude (top). Poles that remain stable in frequency and damping across increasing model order (green) are identified as physical; unstable or new poles at a given order (red crosses) are discarded. Five modes were identified and manually confirmed from this diagram for the specimen shown.

- **S2 – Modal Parameter Uncertainty.** A sustained long-run acquisition at constant low amplitude (approximately 1000 s) was segmented into 60-second macro-segments, and the natural frequency and damping ratio were tracked across segments using a circle-fit SDOF estimator seeded from the S1 poles. The resulting time series were analysed using the overlapping Allan deviation (OADEV) to identify the dominant noise regime and the optimal averaging time for stable modal parameter estimates (Riley, 2008). The per-mode scatter this produces is shown in Figure 3.13, where σ_f ranges from sub-Hz for the lowest, well-separated mode to several Hz for the closely-spaced higher modes. By design, the frequency standard deviation σ_f from S2 was intended to enter directly into the anomaly detection threshold calibration: a change would need to exceed several σ_f before being attributed to structural damage rather than measurement noise. This shared, Group-B-derived threshold was not, in the end, carried forward into the final campaign: Group C1 anomaly detection thresholds are instead derived directly from each specimen's own five-block Phase 3 baseline (Section 5.1 of Chapter 5), which made a shared S2-derived noise constant unnecessary once every Group C1 specimen's own baseline was available. S2 also provides a stationarity check, confirming that the healthy specimen's modal parameters are stationary over the duration of a Phase 3 baseline run, which is a necessary condition for the multivariate Gaussian baseline model underlying the Hotelling T^2 detector to be valid; this role is

unaffected by the change above.

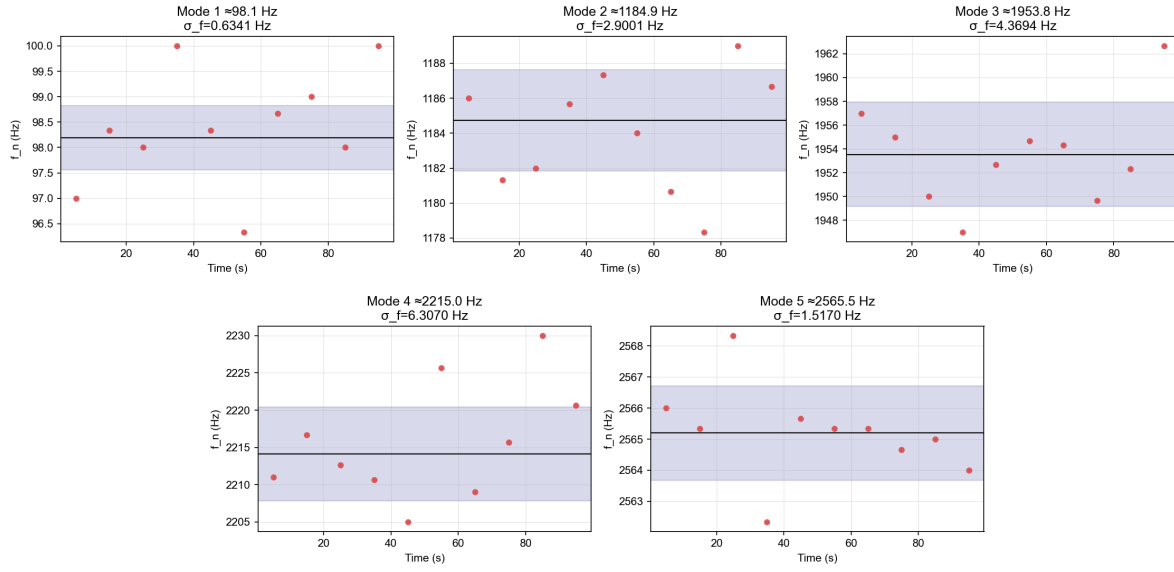


Figure 3.13: S2 modal parameter uncertainty: tracked natural frequency f_n for each of the five identified modes across repeated short-window estimates, with the mean (solid line) and $\pm\sigma_f$ band (shaded) shown per mode. Frequency uncertainty σ_f varies by mode, from sub-Hz for the lowest, well-separated mode to several Hz for closely-spaced higher modes, illustrating why the anomaly detection threshold must be calibrated per mode rather than applied as a single global value.

- **S3 – Mode Shape Scan.** A scanning LDV measurement across the full grid of scan points was performed under broadband excitation, with complex residues estimated at each scan position using the S1 poles as fixed parameters. The assembled mode shape vectors were used to compute the baseline modal curvature profiles and the reference MAC values.
- **S4 – Nonlinearity Validation.** The specimen was driven at the intended Phase 4 fatigue excitation amplitude and the Harmonic Distortion Index (HDI, Equation 2.13 of the Literature Review) was computed from the LDV velocity power spectral density, with $K = 5$ harmonics and drive frequency $f_{\text{drive}} = f_{n,0}$ established by S1. A specimen response is considered linear when $\text{HDI} < 0.05$, i.e. when super-harmonic power remains below 5% of the fundamental. S4 identified the onset of measurable nonlinearity at a drive level of approximately 0.3 V, the super-harmonic content underlying this result being shown in Figure 3.14, and confirmed that the broadband scanning excitation used throughout Phase 3 and Phase 4 (well below this level) remains within the linear regime, validating the Euler-Bernoulli linearity assumption underlying the modal analysis pipeline for every FRF-based feature in this thesis. This validation is particularly important for the 1 mm thickness: geometric nonlinearity can appear at lower amplitudes than material strain limits would suggest for beams of this slenderness (Ewins, 2000). The Phase 4 *harmonic loading* level, by contrast, was deliberately set above the fatigue limit and, in consequence, above this same nonlinearity onset: the specimen's response during the harmonic loading step of a Phase 4 block is therefore expected to be mildly nonlinear by design, even though the broadband scan that follows it remains linear. This distinction, and its implications for the harmonic-stage features (Section 4.1.6), is discussed further in Chapter 6.

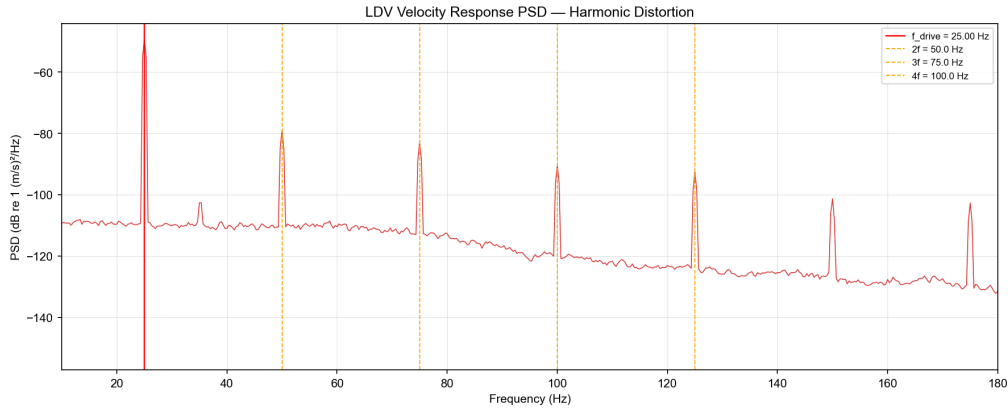


Figure 3.14: S4 nonlinearity validation: LDV velocity response PSD at the drive frequency, showing distinct super-harmonic peaks at $2f$, $3f$, $4f$, and beyond, well above the noise floor. The presence and relative amplitude of these super-harmonics is what the HDI (Equation 2.13) quantifies; their persistence across several harmonics at this drive level is the basis for the reported nonlinearity onset.

- **S5 – EMA versus OMA Consistency.** Experimental Modal Analysis (EMA), which uses the force channel to form the H_1 FRF and identify f_n , ζ , and m_{eff} , was compared against Operational Modal Analysis (OMA), which uses only the LDV velocity response without force normalisation to compute a transmissibility-based modal estimate (Allemang & Avitabile, 2022). This comparison was originally intended to compare the LDV performance with the quantum sensor, but after the change in topic, its main goal is to verify the difference between EMA and OMA on the undamaged specimen before Phase 4 data.

The key outputs of the S1–S5 battery that propagate directly into the fatigue protocol are summarised in Table 3.4.

Table 3.4: Key outputs of the pre-campaign characterisation (S1–S5) and their role in the fatigue test protocol and analysis pipeline.

Test	Output	Role in Phase 3/4 and analysis pipeline
S1	$f_{n,0}$, ζ , m_{eff} ; modal seeds	Drive frequency for Phase 4; seeds for mode tracker (Layer 0)
S2	σ_f , σ_ζ ; optimal averaging time	Anomaly detection threshold calibration (Layer 0); stationarity confirmation
S3	Baseline mode shapes ϕ_0 ; curvature profiles	MAC, CDI, and MSE-TVD reference (Layer 0 spatial features)
S4	HDI at Phase 4 amplitude	Linear regime confirmation; upper scan amplitude bound
S5	EMA/OMA discrepancy; cross-MAC	Validates OMA-mode tracking reliability in Phase 3/4

3.4.3. Load Configuration Study

The five Group B specimens (Section 3.2.3) were used for an iterative load configuration study, rather than a fixed one-configuration-per-specimen test matrix. The right combination of rig stiffness, sample placement, and excitation level was not known at the outset, so the study proceeded in stages: first refining the rig and clamping setup (Section 3.3.1), then the specimen placement and shaker attachment position, and finally the excitation level. Because each stage's outcome informed the next, some specimens were tested under a single configuration while others were tested under two configurations in sequence, once the first configuration's outcome indicated a further adjustment was needed. This same study is also where the final notch geometry itself was settled: an initially planned rectangular-cut notch proved unreliable to manufacture with consistent geometry, and was replaced with the V-notch profile described in Section 3.2.2 once this became apparent.

A further consideration specific to this study is that the microshaker is only practically controllable by drive *voltage*, not by directly commanded force or displacement. Under harmonic excitation in the mass-controlled regime, the resulting displacement amplitude at a given drive voltage scales approximately as $X \approx F/(m(2\pi f)^2)$: at fixed voltage (and therefore approximately fixed force), higher drive frequency directly reduces the achievable displacement. This is why the 30 Hz drive frequency was not simply the lowest convenient choice within the shaker's force budget (Section 3.2.2), but specifically the highest frequency, among those tested, still capable of producing a displacement amplitude large enough to initiate and sustain fatigue damage at the available drive voltage; part of the Group B study's purpose was locating this upper frequency bound empirically.

The selection criterion for the final configuration was threefold: fracture had to occur consistently at the notch location rather than at the clamp or the load cell adhesive bond; the total cycles to fracture had to fall within a range corresponding to between one and twenty Phase 4 blocks of 7,500 cycles, giving adequate temporal resolution for damage tracking without excessively long per-specimen tests; and the force coherence across the broadband scan had to remain above an average of 0.85 across the frequency range containing the first three bending modes, ensuring adequate FRF estimation quality at every block. Figure 3.15 shows the coherence achieved at a candidate configuration against this threshold.

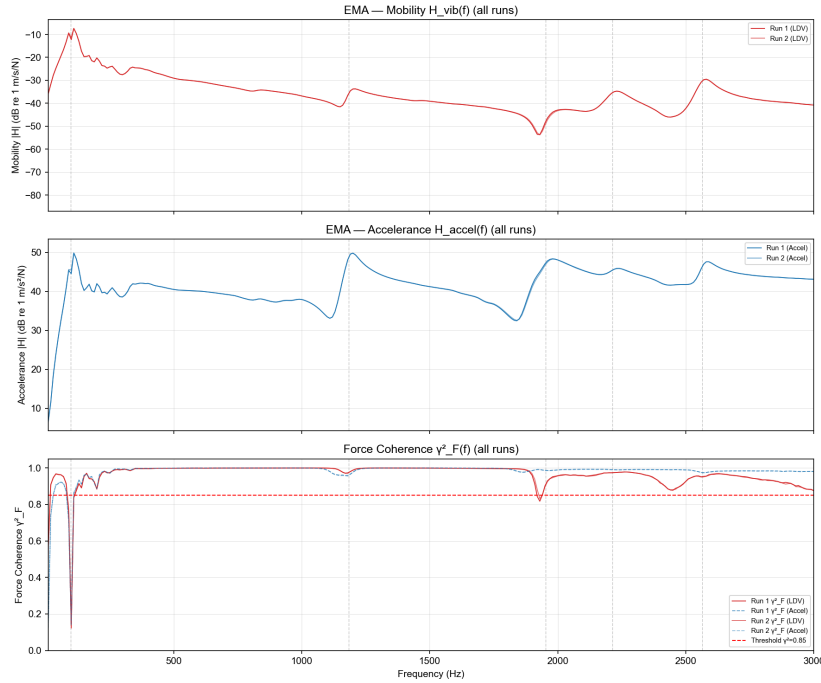


Figure 3.15: Load configuration study: EMA mobility (top) and accelerance (middle) magnitude, and force coherence γ^2_{FF} (bottom), for repeated runs at a candidate configuration. Apart from the first mode, the 0.85 coherence threshold (dashed) is met across the first several bending modes for both the LDV and reference accelerometer channels, confirming adequate FRF estimation quality at this configuration.

The configuration satisfying all three criteria, with the excitation level expressed as a target drive voltage rather than a directly calibrated force in newtons (Section 4.5 discusses the pending force/stress calibration), was adopted for all ten Group C specimens. The shaker attachment position that emerged from this process sits beyond the notch, toward the free end (Section 3.3), rather than at any particular displacement- or stress-optimal point: it was arrived at empirically as the configuration that reliably produced notch-driven fracture within the target cycle range, not derived from a closed-form optimality argument.

The Group B run-to-failure data are not included in the indicator sensitivity or prognostic horizon analyses, as the excitation configurations differ between specimens and across the Group B population.

They are retained as contextual records and referenced in the Discussion where relevant to the interpretation of fatigue life scatter in the final campaign.

3.5. Fatigue Test Protocol

Each of the ten Group C specimens was subjected to an identical two-phase Phase 3/Phase 4 test protocol up to the point of Phase 4 interruption or fracture. Phase 3 establishes the healthy baseline statistical model against which Phase 4 damage progression is measured. Phase 4 is the fatigue loading phase that, for the eight Group C1 specimens, runs until fracture, and, for the two Group C2 specimens, runs until the live-monitoring interruption described in Section 3.5.5. The two phases share the same block structure and measurement cadence by design: this ensures that the feature vectors extracted from Phase 3 and Phase 4 are directly comparable, that the Phase 3 data can serve as the training set for the anomaly detection model without any normalisation for protocol differences, and that the first Phase 4 block can be assessed immediately against the baseline without any transition artefact. An overview of the complete per-specimen protocol is shown in Figure 3.16.

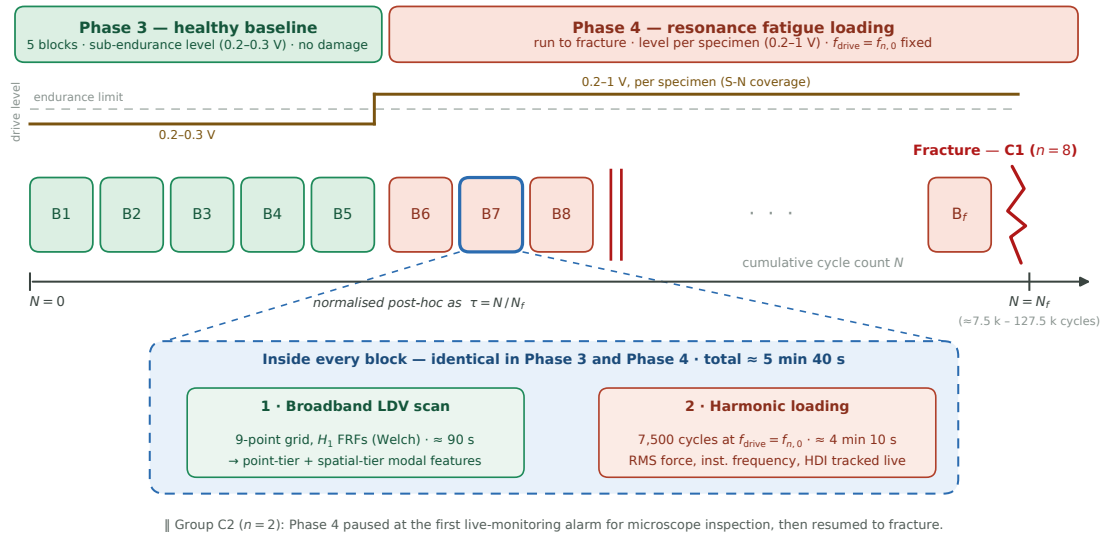


Figure 3.16: Per-specimen test protocol. Phase 3 (healthy baseline, 5 blocks at sub-endurance-limit amplitude) and Phase 4 (fatigue loading at target amplitude, run to fracture; block count varies by specimen) share an identical block structure. The normalised cycle fraction $\tau = N/N_f$ is computed post-hoc once N_f is known from fracture.

3.5.1. Block Structure

The fundamental unit of both phases is the *block*: a fixed-duration fatigue loading stage followed immediately by a broadband modal characterisation scan. Each block proceeds in two sequential steps.

Step 1 – Broadband scanning LDV. Immediately before the harmonic loading stage, the shaker amplitude is reduced to the broadband scan level (confirmed to be in the linear regime by S4 during pre-campaign characterisation) and the PSV-500 scans the full grid of measurement points along the beam neutral axis sequentially. At each scan point, the H_1 FRF between the LDV velocity and the load cell force is computed from a Welch-averaged cross-spectral estimate. The complete multi-point FRF data set provides both the point-tier features (extracted at the reference scan point nearest to the drive point) and the spatial-tier features (assembled from the full mode shape vectors). The scan constitutes the sole modal characterisation measurement of the block; there is no separate single-point sweep. After the final scan point, the data are written to disk, and the pipeline processes the block before the next harmonic loading stage begins.

Step 2 – Harmonic loading. The shaker drives the specimen at the fixed frequency $f_{\text{drive}} = f_{n,0}$, where $f_{n,0}$ is the initial natural frequency of the specimen as identified by S1 during pre-campaign characterisation. The drive frequency is held constant throughout all blocks of both Phase 3 and Phase 4

for that specimen. The harmonic excitation runs for $N_{\text{block}} = 7,500$ cycles per block. During this step, the LDV measures the single-point velocity response at a fixed reference point on the beam, and the load cell records force and shaker-side acceleration continuously. From these signals, three sub-block features are extracted without interrupting the fatigue loading: the instantaneous RMS force (a shaker drift monitor), the instantaneous resonance frequency tracked from the peak of the running PSD, and the Harmonic Distortion Index computed from the full-block Welch PSD.

This two-step sequence is illustrated schematically in Figure 3.17. The block duration is dominated by the harmonic loading time: at 30 Hz, 7,500 cycles correspond to approximately 4 minutes 10 seconds of continuous loading. The broadband scan adds 10 seconds per scan point across the 9-point grid (Section 3.3.2), approximately 90 seconds, giving a total per-block duration of approximately 5 minutes 40 seconds.

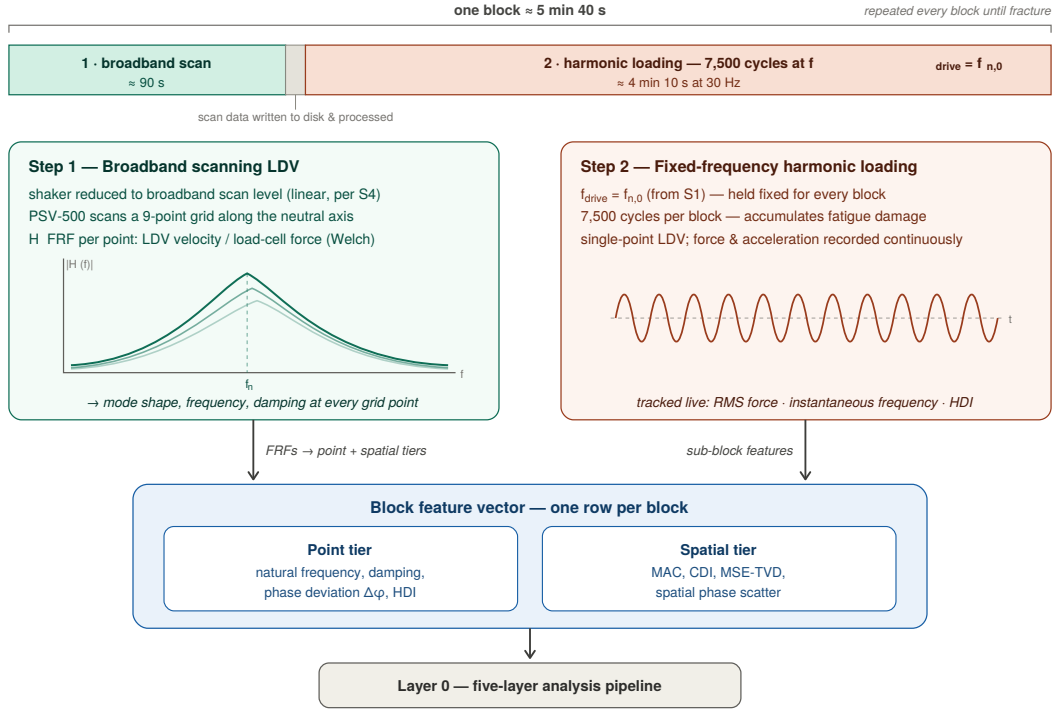


Figure 3.17: Structure of a single fatigue block. The harmonic loading step accumulates damage and provides instantaneous frequency, amplitude, and HDI features; the broadband scanning step extracts the full modal feature vector (point tier and spatial tier) that enters the five-layer analysis pipeline.

3.5.2. Phase 3 - Healthy Baseline

Phase 3 consists of five consecutive blocks executed at a drive level below the fatigue limit of Al 6061-T6 for that specimen, so that no intended damage accumulation occurs during this phase. Rather than a single target shared by every specimen, the Phase 3 level was set individually for each specimen (in the range 0.2–0.3 V drive voltage across the Group C1 population), remaining constant throughout that specimen's Phase 3 blocks. This specimen-by-specimen level is recorded alongside each specimen's results. A stress conversion is not reported here since the force/stress calibration described in Section 4.5 remained pending. The sub-0.3 V range was, however, confirmed by test S4 to sit below the nonlinearity onset identified for this geometry, and is well below the endurance limit $\sigma_{\text{end}} = 97$ MPa expected for this alloy at the corresponding force level. The 5-block Phase 3 run constitutes approximately 20 minutes of continuous operation at 30 Hz, during which the measurement system has warmed up to thermal equilibrium, and any initial bedding-in of the specimen-fixture interface has settled.

The primary purpose of Phase 3 is to populate the healthy feature matrix $\mathbf{X}_0 \in \mathbb{R}^{5 \times p}$ (five blocks, p features) from which the anomaly detection baseline model is fitted at the conclusion of Phase 3. Stationarity of all features across the five blocks is a necessary condition for the multivariate Gaussian

baseline model. Any systematic drift observed in the Phase 3 diagnostic plots, particularly in the phase quadrature deviation $\Delta\phi$ or the HDI, is treated as evidence of fixture instability and triggers an investigation before Phase 4 begins. The Phase 3 data also populate the reference mode shapes ϕ_0 that serve as the MAC and CDI baseline for Phase 4 tracking: because Phase 3 uses the identical block structure as Phase 4, these reference shapes are measured under the same scan conditions as all subsequent Phase 4 scans, eliminating any protocol-induced shape inconsistency.

3.5.3. Phase 4 - Resonance Fatigue Loading

Phase 4 begins immediately after Phase 3 completes, without remounting the specimen or interrupting the acquisition session. The shaker level is stepped up from the Phase 3 level to the Phase 4 fatigue target for that specimen. Unlike Phase 3, the Phase 4 target level was deliberately varied across the Group C1 population (drive levels in the range 0.2–1 V, chosen per specimen), and, for the earliest specimens, the shaker attachment position was also varied (Section 3.3.1), rather than either being held at one fixed value, for two reasons: first, testing at several distinct stress levels, whether reached via drive voltage or via attachment position, was necessary to populate more than one point on the Basquin S-N fit of Section 4.5; second, holding the level fixed across every specimen risked the Layer 0.5 autoencoder (Section 4.2) learning to key on a single, constant excitation amplitude rather than on the damage-relevant spectral shape, which a range of training levels helps to guard against. Within a given specimen's Phase 4 run, both the level and the attachment position are held constant. All Phase 4 levels exceed the endurance limit and lie above the nonlinearity onset identified by S4 (Section 3.4.2). The drive frequency remains fixed at $f_{n,0}$ throughout Phase 4 regardless of the amplitude level in use.

Phase 4 proceeds as a sequence of blocks identical in structure (though not, as discussed above, in excitation level) to Phase 3. For the eight Group C1 specimens, the test is continued without interruption until physical fracture occurs, observed as a sudden loss of load cell signal and a sharp amplitude drop in the LDV response. No artificial termination criterion based on damage index threshold is applied. The block-level cycle count at fracture is a coarse frame only: the within-block fracture-detection logic (Section 4.6) resolves the exact cycle of fracture inside that block, so two specimens that fracture nominally "in the same block" can still be distinguished and ordered by their precise N_f . Across the eight Group C1 specimens, N_f ranged from approximately 7,500 to 127,500 cycles, reflecting the natural inter-specimen scatter in crack initiation timing and propagation rate for nominally identical specimens under deliberately varied loading levels. For the two Group C2 specimens, Phase 4 instead follows the interrupted protocol detailed in Section 3.5.5: loading stops at the first live-monitoring alarm, the specimen is removed for microscope re-inspection, and loading then resumes to final fracture. Because the interruption point is determined by the detection logic rather than fixed in advance, the cumulative cycle count at interruption is not directly comparable across the two Group C2 specimens, nor with the Group C1 N_f range. The eventual fracture cycle count for each Group C2 specimen is recorded but is reported individually rather than pooled into the Group C1 statistics.

The fixed-frequency choice and its rationale were established in the chapter introduction (Section 2.3.2 of the Literature Review). Its practical consequence at the block level is that the response amplitude decreases gradually as the specimen's natural frequency drifts below $f_{n,0}$ during Phase 4: the applied stress amplitude is not constant, decreasing as the Dynamic Amplification Factor (DAF) falls with increasing detuning. The RMS force is monitored at each block to distinguish genuine structural changes from shaker drift, but no correction is applied to the excitation level. The gradually decreasing stress amplitude is treated as an observable feature of the damage process (it contributes to the compliance feature) and its implications for the interpretation of damage rate and for comparison with constant-amplitude S-N data are addressed in the Discussion (Section 6.5).

3.5.4. Data Storage

All raw time-series data and extracted features for both Phase 3 and Phase 4 are stored in a single HDF5 file per specimen, named by specimen identifier. The schema is designed for incremental append: each block appends one new row to the point-feature dataset and, where a scan was performed, one row to the spatial-feature dataset, without rewriting earlier data. This allows the analysis pipeline to be run after each block during live testing, providing near-real-time damage index and RUL updates, or run in batch after the full campaign has completed.

The top-level schema is organised into two groups – /baseline/ (Phase 3 blocks and the fitted baseline model) and /fatigue/ (Phase 4 blocks, anomaly scores, damage indices, and RUL estimates) – with raw block data stored under `block_{BB}`/ subgroups within each. The scan grid coordinates are stored alongside the spatial feature data, ensuring that all spatial plots are reproducible from the HDF5 file alone without reference to external configuration. A summary of the key HDF5 datasets and their roles in the analysis pipeline is given in Table 3.5.

Table 3.5: Key datasets in the per-specimen HDF5 file and their roles in the analysis pipeline.

Dataset path	Dimensions	Content and role
/baseline/point_features	$5 \times p_1$	Point-tier features from Phase 3 blocks; input to baseline model fitting
/baseline/spatial_features	$5 \times p_2$	Spatial-tier features from Phase 3 scans; reference MAC and curvature
/baseline_model/	arrays	Fitted μ , Σ^{-1} , T^2 threshold, Isolation Forest model, CUSUM parameters
/fatigue/point_features	$N_b \times p_1$	Growing point-tier feature matrix; one row per Phase 4 block
/fatigue/spatial_features	$N_b \times p_2$	Growing spatial-tier feature matrix
/fatigue/anomaly_scores	$N_b \times 4$	T^2 , Isolation Forest score, CUSUM value, damage flag (Layer 0)
/fatigue/cae_error	$N_s \times 2$	HI1 (reconstruction MSE), HI2 (latent deviation) per scan block (Layer 0.5; absent if Layer 0.5 was not run)
/fatigue/damage_indices	$N_b \times 6$	D_{T^2} , D_{freq} , D_{PC1} , D_{comp} , D_{CAE} , $D_{\text{CAE,latent}}$ (Layer 2; the final two columns are all-zero with a logged warning if /fatigue/cae_error is absent)
/fatigue/rul_estimates	$N_b \times 3$	RUL (trend), RUL (GP), 95% CI width (Layer 3)
/fatigue/scan_coordinates	$N_s \times 2$	Scan point positions $[x_i, y_i]$ in mm

3.5.5. Independent Validation via Optical Fractography

The Layer 0–3 analysis pipeline infers damage exclusively from vibration response: at no point does it observe a crack directly. The two Group C2 specimens (Section 3.2.3) were set aside to provide an independent check on this inference chain, using periodic optical microscopy as a vibration-free ground truth against which the pipeline’s own detection logic could be evaluated. The validation protocol for each Group C2 specimen proceeds in four stages.

- **Stage 1 – Pre-test characterisation.** Before Phase 4 loading began, the notch root and the expected crack initiation site were imaged under an optical microscope at a fixed magnification and stage position, established once per specimen and reused for every subsequent image so that successive images are directly comparable. This pre-test image constitutes the specimen’s undamaged reference state, against which any later surface change is assessed qualitatively rather than through an automated measurement, for the reasons discussed in Section 2.7.5 of the Literature Review.
- **Stage 2 – Phase 4 loading with live monitoring.** The specimen was then loaded under the identical fixed-frequency Phase 4 protocol used for the Group C1 specimens, with the live-monitoring system (Section 4.6) watching the specimen’s data folder in real time as each block completed.

Loading continued, exactly as for a Group C1 specimen, until the live monitor raised its first damage alarm, that is, until any one of the Hotelling T^2 , HDI, or CAE-based health-indicator tests exceeded its Phase 3-derived threshold for a completed block.

- *Stage 3 – Interruption and re-inspection.* At the first damage alarm, loading was stopped immediately and the specimen was returned to the microscope at the same magnification and stage position used in Stage 1. The resulting image provides a second, vibration-independent snapshot of the specimen's physical state at the exact life fraction the live monitor flagged as anomalous, allowing a direct qualitative comparison: is a crack, or an early-stage surface change consistent with crack initiation, visible at the notch root at the point where the statistical or learned indicator first crossed its threshold?
- *Stage 4 – Final fracture and fracture-surface comparison.* Following re-inspection, the specimen was remounted and Phase 4 loading resumed under the same fixed-frequency protocol, continuing until fatigue fracture occurred. Breaking the specimen by hand was considered and rejected: a manually induced overload fracture would not be comparable with the Group C1 surfaces, which is precisely the comparison this stage exists to make. The resulting fracture surface was compared, under the same microscope, against the fracture surfaces of the eight Group C1 specimens run to fracture without interruption, to establish whether the detected damage was compatible with fatigue-related phenomena.

Consistent with the qualitative role of fractography established in the Literature Review, this validation does not produce a quantitative crack-length time series. It produces a paired sequence of images, per specimen, that supports or challenges the detection-timing claims made by the statistical and learned indicators elsewhere in the pipeline. The two Group C2 specimens are therefore reported and discussed individually in the Results and Discussion chapters rather than pooled into the population-level statistics drawn from Group C1.

4

Analysis Pipeline Methodology

This chapter describes the analysis methodology applied to the block-by-block data produced by the experimental campaign of Chapter 3. Like Chapter 3, it is methodological rather than results-bearing: it establishes how each research question is addressed computationally, but the actual outcomes, comparisons, and interpretation are deferred to the Results and Discussion chapters.

The chapter's structure follows the logical shape of the analysis itself, which is not a single sequential pipeline but a fork followed by two independent side investigations. Every block of raw signal data can be reduced to a set of candidate damage indicators in two genuinely different ways: by computing a hand-designed feature vector grounded in the modal indicators surveyed in the Literature Review (Section 3.1), or by learning a representation of the healthy state directly from the raw scan data without specifying in advance what to look for. These two tracks are not sequential stages of one pipeline; they are alternative answers to the same question, computed independently and compared directly. Both converge on a shared problem: given a pile of candidate indicators from either or both tracks, which ones actually track damage, and how should they be combined into a single, continuous damage index? That combined index then feeds directly into an estimate of remaining useful life. Two further investigations sit outside this chain entirely. Characterising the campaign against classical stress-life (S-N) testing depends only on the raw force and cycle-count records, not on any of the indicators or damage index constructed above. Live monitoring and the fractography-based validation of Chapter 3 address a different question again: does any of this work in real time, and does it correspond to a physically real, observable event? Figure 4.1 shows this structure, and Table 4.1 maps each section onto the research question it addresses.

Table 4.1: Mapping of chapter sections onto the research questions they address.

Section	Core question answered	RQ
4.1 From Raw Signal to Modal Feature Vector	How does the vibration response evolve as damage accumulates?	R1
4.2 An Alternative Construction: Learned Health Indicators	Can a learned, feature-free representation match hand-engineered sensitivity?	R2 (comparative)
4.3 From Many Indicators to One Damage Index	Which indicators are most sensitive, and how are they combined into $D(\tau)$?	R2
4.4 Estimating Remaining Useful Life	How much life remains, and from what point is that estimate reliable?	R3
4.5 S-N Characterisation and the Resonance-Testing Context	How does this campaign's fatigue life compare to classical stress-life testing?	R4
4.6 Live Monitoring and Independent Validation	Does detection work in real time, and does it correspond to a physically real event?	Validation

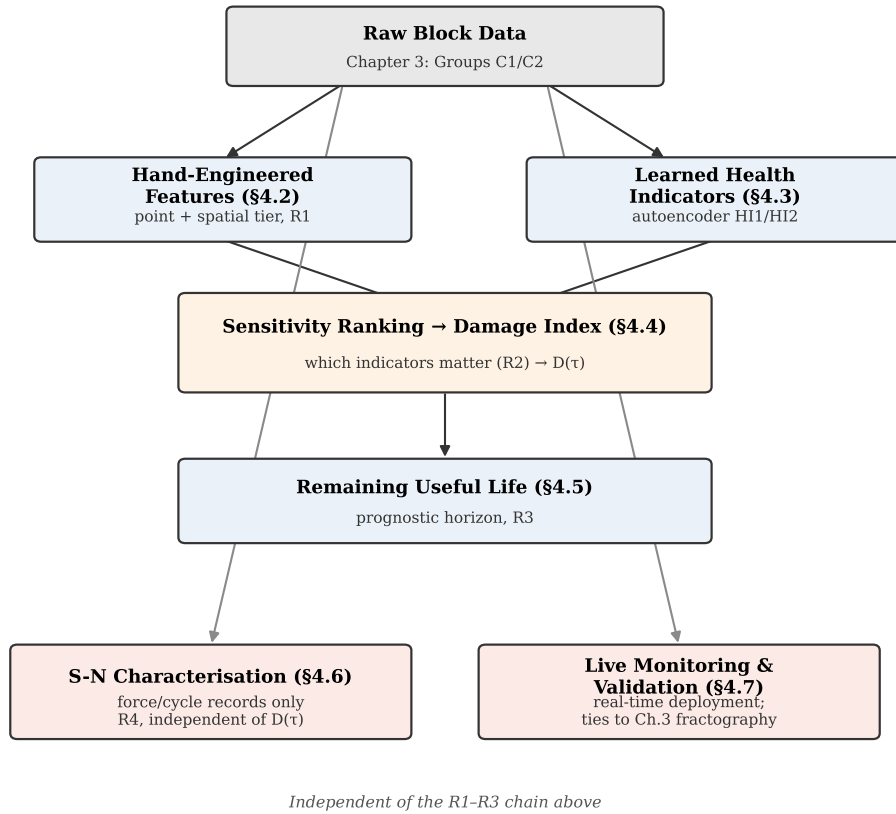


Figure 4.1: Structure of the analysis methodology described in this chapter. Raw block data feeds two independent indicator-construction tracks (Sections 4.1 and 4.2), which converge on a shared sensitivity-ranking and damage-index step (Section 4.3) feeding remaining-useful-life estimation (Section 4.4). S-N characterisation (Section 4.5) and live monitoring/validation (Section 4.6) are independent of this chain, requiring only the raw force/cycle records and the real-time data stream respectively.

4.1. Signal Processing and Modal Feature Extraction (Layer 0)

Layer 0 converts the raw time-domain signals recorded during each fatigue block into a fixed-length feature vector of physically interpretable modal parameters. It answers the question: what is the current dynamic state of the specimen? Most subsequent analysis layers operate on this feature vector rather than on raw signals, ensuring that the machine-learning components of the pipeline are grounded in structural mechanics throughout. The processing chain is identical for Phase 3 and Phase 4 blocks, preserving direct comparability between the healthy baseline and the damage-accumulation records.

4.1.1. Indicator Selection Rationale

The Literature Review (Section 2.5) surveyed the global and spatial damage indicators reported in the wider vibration-based fatigue monitoring literature: five scalar global indicators (frequency shift, damping ratio, compliance, phase deviation, HDI) and four spatial indicators (MAC, CDI, MSE-TVD, spatial phase scatter). This thesis computes all nine, rather than a pre-selected subset, for two reasons specific to the present experimental design. First, the multi-indicator comparison itself is the object of Research Question R2: narrowing the feature set in advance would pre-empt the very ranking the thesis sets out to produce. Second, the fixed-frequency protocol adopted here (Section 3) changes the expected relative performance of several indicators in ways the tracked-frequency literature cannot predict: phase deviation in particular is expected, on the physical grounds established in Section 2.4, to be unusually informative under a fixed drive frequency precisely because the protocol does not correct for detuning, so excluding it on the strength of its more modest reported performance under tracked-

frequency protocols elsewhere in the literature would risk discarding the indicator most likely to benefit from this thesis's specific experimental choice.

Two indicators from the literature survey are adapted rather than used unchanged. The phase-deviation feature is generalised from the single-structural-point formulation of Di Maio et al. (Di Maio et al., 2018) to a phase difference between the LDV response and the shaker-body acceleration (Section 3.3.1), since the present fixture does not provide a convenient second structural response point; this adaptation is described in full, together with the resulting feature definition, in Section 4.1.6. The spatial phase standard deviation is likewise a direct extension of the same underlying concept to the scanning measurement, made possible because the protocol performs a full spatial scan after every block rather than at a coarser cadence, itself a deliberate departure from the global/spatial acquisition-cost asymmetry discussed in Section 2.5, adopted here because the additional scan time was judged affordable within the fatigue block budget and removes scan cadence as a confound when comparing global and spatial indicator sensitivity for Research Question R2. The complete feature vector, with the precise definition and computation of every indicator selected here, is given in Section 4.1.6.

4.1.2. Signal Pre-processing

All raw time-series data are acquired at $f_s = 10,000$ Hz (Section 3.3.3). No further decimation is applied. At this rate, the Nyquist frequency of 5 kHz already sits comfortably above the third bending mode expected for this geometry (below 1500 Hz), so all spectral analysis in this thesis operates directly on the acquired signal.

A linear trend is removed from each signal segment before windowing to suppress low-frequency drift that would otherwise leak into the spectral estimates. No additional pre-filtering is applied. The Hann windowing used in the Welch estimator provides adequate sidelobe suppression for the modal frequency range of interest.

4.1.3. FRF Estimation and Modal Identification

Broadband scan blocks

For each scan point x_i , the H_1 FRF between the LDV velocity and the load cell force is estimated via Welch's method (Section 2.4 of Chapter 2), using overlapping Hann-windowed segments of 1 s duration with 50% overlap. Force coherence $\gamma^2(f)$ is computed alongside the FRF as a quality gate; scan-point measurements with coherence below 0.85 at resonance are flagged but retained in the dataset so that the coherence degradation itself can serve as a damage feature. The 1 s segment length gives a spectral resolution of $\Delta f = 1$ Hz regardless of sample rate, placing the frequency bins well below the expected first-mode frequency shift of order 0.1–1% over the specimen's fatigue life (0.03 – 0.30 Hz at $f_{n,0} \approx 30$ Hz), resolvable by the sub-bin interpolation applied in pole extraction.

Per-mode identification

Natural frequencies and damping ratios are extracted from the H_1 FRF using one of the modal identification methods surveyed in Section 2.6.1 of Chapter 2; the pipeline's default is the SDOF circle-fit method, fast and well suited to the well-separated, lightly-damped modes it assumes. For this campaign specifically, however, N4SID (the time-domain subspace alternative) was the method actually used to extract the modal parameters reported in this thesis, not circle-fit. The switch was necessary rather than a preference: circle-fit's isolated-resonance assumption was violated by the elevated measurement noise and closely-spaced, highly-damped modes encountered in this campaign's FRF data (Section 4.1.4 below), conditions under which a frequency-domain, per-mode fit degrades sharply while N4SID's time-domain state-space realisation does not share the same failure mode. The full mechanistic explanation for why N4SID was better suited to these specific conditions, and the evidence for the noise and close-mode-spacing conditions themselves, is given in Chapter 6 (Section 6.1.1) rather than repeated here; this subsection records only the identification method actually used and the empirical reason the default was replaced. The search band for each mode is initialised from the S1 baseline identification and updated by the mode tracker at each block, using the same MAC-based pole disambiguation described in Section 4.1.4 regardless of which identification method underlies it.

Harmonic loading blocks

During the harmonic excitation stage of each block, no broadband FRF is estimated. Instead, the instantaneous resonance frequency and response amplitude are tracked by locating the peak of the running Welch power spectral density of the LDV velocity signal. The HDI is computed from the full-block PSD as defined in Equation 2.13. These sub-block features provide a higher-cadence record of the frequency drift *during* loading, complementing the broadband scan measurements taken between loading stages.

4.1.4. Mode Tracking Across Blocks

Fatigue damage causes a monotonic downward drift in the natural frequencies of the specimen. A fixed frequency search band would eventually lose track of a mode as the resonance migrates out of the window. The ModeTracker maintains a running estimate of the centre frequency for each tracked mode using an exponential moving average updated at every block:

$$f_{\text{tracked}}(t) = \alpha f_{\text{measured}}(t) + (1 - \alpha) f_{\text{tracked}}(t - 1) \quad (4.1)$$

The smoothing factor in Equation 4.1 is set to $\alpha = 0.05$. This value is deliberately conservative: for a metallic beam under constant-amplitude resonance fatigue, the true natural frequency shifts by well under 0.5% per block, so a heavier smoothing factor is appropriate and a lighter one risks the tracker reacting to single-block noise as though it were a genuine shift. A jump-rejection filter provides a second line of defence: if the extracted frequency deviates by more than 3% from the running estimate, the measurement is flagged, excluded from the EMA update, and the tracker retains its previous value. This rejection can be disabled if the tracker becomes permanently stuck on a stale estimate, which can occur if the initial frequency guess is wrong or if a genuine drift proceeds gradually past the rejection threshold over many blocks before the tracker catches up.

This EMA-and-rejection scheme alone proved insufficient in practice. Both mechanisms operate purely on frequency proximity and amplitude, with no way to distinguish a genuine structural resonance from a transient noise peak or a neighbouring mode leaking into the search band other than "closest to the tracked value" or "tallest in the band", an ambiguity that caused the tracked frequency to drift noticeably even on healthy blocks, before either mechanism below was added. The first is an automatic, always-on improvement: *frame-by-frame MAC-based pole tracking* disambiguates competing candidate poles within a mode's search band by comparing each candidate's mode shape to the previous block's accepted shape, rather than relying on frequency proximity or amplitude alone, since a spurious or neighbouring-mode pole is far less likely to also match the expected mode shape closely. The second is an optional manual override: a *pole-selection review* interface lets the researcher inspect the stabilisation diagram for any specific block and lock in a chosen pole by hand, for cases where the automatic disambiguation itself is uncertain. The two are complementary rather than alternatives: the first handles the overwhelming majority of blocks without intervention, the second exists for the exceptional case.

The tracker simultaneously maintains a time series of all tracked modal parameters ($f_n(t)$, $\zeta(t)$, $|H(f_n)|^2(t)$, $\gamma^2(f_n)(t)$), which collectively constitute the primary output of Layer 0 and the answer to Research Question R1.

4.1.5. Manual Pole-Selection Review and Modal Identification Quality Checks

The pole-selection review interface introduced above lets the researcher inspect the N4SID stabilisation diagram for any specific specimen, phase, and block, and lock in a chosen pole by hand where the automatic MAC-based disambiguation is judged uncertain. The diagram plots candidate poles across a sweep of model orders against the H_1 FRF magnitude, classified as stable in both frequency and damping, frequency-stable only, or unstable/noise, with the automatic selection for that block marked for direct comparison; the researcher assigns a clicked pole to a tracked mode and confirms the block, after which that mode's parameters are locked at the manually assigned value for downstream processing rather than the automatic estimate. This review was used selectively, on specimens and blocks where the automatic tracker's estimate looked doubtful on inspection, rather than as a parallel full re-analysis of every block.

A second, independent diagnostic, the cross-mode Modal Assurance Criterion matrix, is available for checking modal identification quality directly rather than relying on the damage indicators to reveal a problem indirectly. Whereas the MAC value used elsewhere in this thesis as a damage indicator (Section 4.1.6) is a single current-vs-baseline diagonal comparison for one mode, the full cross-mode matrix compares every tracked mode against every other tracked mode within the same block. The diagonal reproduces the current-vs-baseline damage indicator; the off-diagonal terms answer a different question, whether two supposedly distinct tracked modes have converged onto similar shapes. A high off-diagonal value (above a threshold of 0.95) indicates such mode mixing, a modal identification problem, rather than evidence that the structure itself has changed, and is flagged visually where it occurs. This matrix is read directly from the mode shapes already stored by Layer 0 and requires no signal reprocessing.

Both diagnostics are reported for the Group C1 campaign in Chapter 5 (Section 5.2), alongside the finding that only two of the three intended modes could be tracked with acceptable stability; the cross-mode matrix in particular is the evidence bearing on whether those two tracked modes correspond to the same physical mode consistently across the eight specimens.

4.1.6. Feature Vector

The Layer 0 feature vector is organised into two tiers according to where the underlying measurements are extracted. The *point tier* contains features evaluated at the reference scan point and from the harmonic loading stage, whereas the *spatial tier* contains features computed from the full LDV scan. Although the two tiers reflect different measurement domains, both are updated after every fatigue block in the present campaign because the broadband scan is performed after every block ($M = 1$).

All damage-oriented features are defined relative to the Phase 3 healthy baseline and follow a common sign convention in which the healthy state is zero or near-zero and increasing damage produces a positive trend where physically appropriate. The complete Layer 0 vector contains 33 scalar inputs; Table 4.2 summarises the underlying feature families and their expansion into the final vector.

Point tier - extracted every block

The point-tier vector $\mathbf{x}_1 \in \mathbb{R}^{p_1}$ combines measurements from the broadband scan at the reference scan point (the point nearest to the shaker attachment) with features extracted during the harmonic loading stage.

For each tracked mode i , the broadband measurements provide the relative natural-frequency shift $\Delta f_{n,i}(t) = \frac{f_{n,i}(0) - f_{n,i}(t)}{f_{n,i}(0)}$, the absolute damping change $\Delta \zeta_i(t) = \zeta_i(t) - \zeta_i(0)$, the FRF peak compliance $|H(f_{n,i})|$, and the coherence $\gamma^2(f_{n,i})$.

The harmonic stage provides three additional damage-sensitive features:

- *Harmonic Distortion Index* (HDI), defined as the ratio of super-harmonic power to fundamental power in the harmonic-loading PSD (Equation 2.13). It is sensitive to the crack-breathing nonlinearity generated during fatigue damage.
- *Phase quadrature deviation* ($\Delta \phi_{\text{diff}}$), defined as the deviation of the phase difference $\phi_{\text{diff}}(f) = \angle H_{\text{LDV}}(f) - \angle H_{\text{shaker}}(f)$ from its Phase 3 baseline mean at the fixed drive frequency $f_{n,0}$. This extends the phase-difference indicator of Di Maio (Di Maio et al., 2018) to the fixture configuration described in Section 3.3.1, in which H_{shaker} represents the FRF of the shaker-body acceleration rather than a second structural response point.
- *Phase difference slope* ($d\phi_{\text{diff}}/df$), obtained by linear regression over a ± 5 Hz window around $f_{n,0}$. The slope is inversely related to modal damping and therefore flattens as damping increases during crack growth. It can consequently provide an indication of damage before a measurable frequency shift occurs.

Two additional harmonic-stage quantities, RMS force and harmonic amplitude, are retained as monitoring features. These primarily track shaker drift rather than structural damage and are therefore excluded from the damage-indicator analysis.

Spatial tier - extracted every block

The spatial-tier vector $\mathbf{x}_2 \in \mathbb{R}^{p_2}$ is derived from the full broadband scan. For each tracked mode i , it contains the Modal Assurance Criterion (MAC_i), which measures the correlation between the current mode shape $\phi_i(t)$ and the Phase 3 reference shape $\phi_{i,0}$ (Equation 2.14). A reduction in MAC indicates distortion of the mode shape associated with local stiffness loss.

The spatial tier also contains the maximum curvature-difference measure, $\max_x \text{CDI}_i$, defined from the absolute curvature difference $|\kappa_i(x, t) - \kappa_{i,0}(x)|$.

Curvature provides greater local sensitivity to damage than the mode-shape amplitude itself because, for a beam in bending, it is directly related to local bending strain and therefore responds strongly to local stiffness loss. The complete CDI field is retained separately for damage localisation.

Curvature is evaluated using the unequal-spacing finite-difference expression

$$\kappa_i = \frac{2 \left(\frac{\phi_{i+1}}{h_2} - \phi_i \left(\frac{1}{h_1} + \frac{1}{h_2} \right) + \frac{\phi_{i-1}}{h_1} \right)}{h_1 + h_2}, \quad (4.2)$$

where $h_1 = x_i - x_{i-1}$ and $h_2 = x_{i+1} - x_i$ in Equation 4.2 are the local scan-point spacings. The general unequal-spacing form is used because individual scan points may be dropped or flagged by the coherence gate (Section 4.1.3), resulting in a locally non-uniform effective grid.

Two further spatial features capture distributed changes across the scan:

- *MSE total variation distance* (MSE-TVD), which measures the total variation distance between the current and baseline fractional modal strain-energy distributions across the scan elements (Equation 2.16). It integrates spatial damage evidence over the beam and is less sensitive to noise at individual scan positions than a single curvature measurement (Liang et al., 2024).
- *Spatial phase standard deviation* (σ_ϕ), defined as the standard deviation of $\phi_{\text{diff}}(f_{n,0}, x_i)$ across the scan points. It is near zero for the healthy specimen and increases when a crack produces a localised anomaly in the spatial phase field. This extends the phase-based indicator of Di Maio (Di Maio et al., 2018; Magi et al., 2016b) into the spatial domain enabled by the scanning LDV.

Expansion to the complete Layer 0 vector

The feature families above are expanded into the 33 scalar inputs used by Layer 0. Per-mode features are instantiated for both tracked modes, while most features are represented in both their absolute and baseline-referenced `delta_` forms. The curvature measure is represented by both its maximum value and its localisation index. The two harmonic-stage monitoring channels, RMS force and harmonic amplitude, complete the point-tier feature set, although they are not used in the damage-index construction. The result is 20 point-tier and 13 spatial-tier features.

Several resulting feature pairs are near-exact complements by construction: a baseline-referenced feature contains essentially the same information as its absolute counterpart with the sign reversed. These pairs were retained deliberately rather than removed a priori, allowing the redundancy-detection procedure in Section 4.3.3 to identify known redundancies directly from the feature matrix.

The resulting feature architecture is summarised in Table 4.2. The table lists feature families rather than every scalar input; the individual 33 features are enumerated in Table 5.3 of Chapter 5.

Table 4.2: Layer 0 feature families, their extraction domain, and the physical mechanism they are intended to capture.

Tier	Feature family	Extraction domain	Primary damage information
Point	$\Delta f_{n,i}$	Reference scan point	Stiffness loss $\rightarrow$ natural-frequency reduction
Point	$\Delta \zeta_i$	Reference scan point	Crack-face rubbing $\rightarrow$ increased energy dissipation
Point	$ H(f_{n,i}) $	Reference scan point	Stiffness loss and damping increase $\rightarrow$ compliance increase
Point	$\gamma^2(f_{n,i})$	Reference scan point	Crack-breathing nonlinearity $\rightarrow$ coherence reduction
Point	HDI	Harmonic loading stage	Breathing crack $\rightarrow$ super-harmonic generation
Point	$\Delta \phi_{\text{diff}}$	Harmonic loading stage	Stiffness loss $\rightarrow$ structural phase-lag change
Point	$d\phi_{\text{diff}}/df$	Harmonic loading stage	Damping increase $\rightarrow$ phase-slope flattening
Spatial	MAC_i	Full LDV scan	Stiffness loss $\rightarrow$ mode-shape distortion
Spatial	$\max_x \text{CDI}_i$	Full LDV scan	Local stiffness loss $\rightarrow$ curvature anomaly
Spatial	MSE-TVD	Full LDV scan	Damage $\rightarrow$ redistribution of modal strain energy
Spatial	σ_ϕ	Full LDV scan	Localised damage $\rightarrow$ spatial phase scatter

4.1.7. Anomaly Detection Baseline (Layer 0 Output)

The anomaly detection component of Layer 0 answers the binary question: does the current feature vector look like the healthy baseline? It operates exclusively on the point-tier features, for which a complete five-block Phase 3 record is available for every specimen.

After all Phase 3 blocks are processed, a multivariate Gaussian baseline model $\mathcal{N}(\mu, \Sigma)$ is fitted to the Phase 3 point-feature matrix $\mathbf{X}_0 \in \mathbb{R}^{5 \times p_1}$. The covariance matrix is regularised as $\Sigma + \lambda \mathbf{I}$ with $\lambda = 10^{-6}$ to prevent numerical instability when features are nearly collinear. Three detectors are then applied to each Phase 4 block in combination (Elouny et al., 2023):

Hotelling T^2 : the squared Mahalanobis distance from the healthy cluster centre, following the multivariate Gaussian baseline formulation of Section 2.8 of the Literature Review,

$$T^2(t) = (\mathbf{x}(t) - \mu)^\top \Sigma^{-1} (\mathbf{x}(t) - \mu) \quad (4.3)$$

In Equation 4.3, μ and Σ are estimated from the five Phase 3 blocks, and the detection threshold is derived from the F -distribution at $\alpha = 0.01$ (Jolliffe & Cadima, 2016).

CUSUM: applied to the T^2 score time series exactly as introduced in Section 2.8,

$$S(t) = \max(0, S(t-1) + T^2(t) - k) \quad (4.4)$$

The reference value in Equation 4.4 is $k = 0.5$, with the chart signalling when $S(t) > h = 5.0$ (Page, 1954).

Isolation Forest: trained on the Phase 3 feature matrix, using the same algorithm described in Section 2.8 (Liu et al., 2008).

The combination rule is: damage is flagged when (T^2 alarm and Isolation Forest alarm) or when CUSUM alarm fires alone. CUSUM alone is treated as sufficient because sustained drift is inherently more reliable evidence than a single instantaneous exceedance. The anomaly detection output is stored alongside the feature matrix in the HDF5 file and used as a detection timestamp for the prognostic horizon analysis in Layer 3.

4.2. Learned Health Indicators (Layer 0.5)

Layer 0.5 was developed after the core four-layer pipeline described above and is integrated as an optional fifth analysis layer, motivated by the exploratory question identified in the Literature Review (Section 2.8.4): whether a learned, feature-free representation of the healthy state offers detection sensitivity that the hand-engineered Layer 0 feature set does not. Unlike Layers 0, 1, 2, and 3, Layer 0.5 operates directly on the raw broadband scan signal rather than on the extracted point- or spatial-tier feature vectors, and is trained exclusively on Phase 3 (healthy) data. The network never observes a damaged specimen during training, which is what makes its output a valid health indicator at inference time on Phase 4 data.

4.2.1. From Raw Scans to Log-FRF Sub-Windows

Each broadband scan signal is processed into a population of overlapping sub-windows so that the network sees many FRF estimates per block rather than a single one. Each scan is split into overlapping windows of length 5 s at a hop of 2.5 s (50% overlap), Hann-tapered, and converted to a single-window H_1 FRF estimate. The resulting spectrum is cropped to $[0, 500]$ Hz, resampled onto 4,000 uniform frequency bins (decoupling the network input size from the raw FFT bin count), and converted to log-magnitude form, $x = 20 \log_{10}(|H| + \varepsilon)$. Finally, each frequency bin is z -scored using the per-bin mean and standard deviation computed once from the pooled Phase 3 training windows and reused unchanged at inference. The 5 s window length gives a frequency resolution of $\Delta f = 0.2$ Hz, fine enough to resolve a 1% damping ratio at 30 Hz (half-power bandwidth ≈ 0.3 Hz); shorter windows degrade this resolution and are not used.

Training pools sub-windows across every specimen processed in a single Layer 0.5 invocation rather than training one network per specimen: a healthy-state manifold that generalises across nominally identical specimens of the same geometry is the more physically appropriate target than a per-specimen model, and this design choice means the trained network itself, not just its outputs, is shared across the Group C1 population.

4.2.2. Network Architecture and Resonance-Weighted Loss

The network is a one-dimensional convolutional autoencoder (CAE), following the general approach of using a reconstruction-error-based autoencoder trained exclusively on healthy-state data for unsupervised structural anomaly detection (Toufigh & Ranjbar, 2024). The encoder reduces the 4,000-bin log-FRF input through three Conv1D + MaxPool + Dropout stages to a 64-dimensional latent bottleneck – a 62.5:1 compression ratio – and the decoder mirrors the encoder to reconstruct the full 4,000-bin spectrum. A 64-number latent code is sufficient because a healthy FRF is a structured signal: one or two resonance peaks over a smooth background, not an arbitrary high-dimensional pattern. Smaller 16- and 32-dimensional bottleneck configurations were also tested, for comparison, and are reported alongside this primary configuration in Chapter 5.

Reconstruction loss is not applied uniformly across the spectrum. A resonance-weighted mean-square error concentrates the network’s training signal on the frequency bands that carry physical information about the structure’s health, rather than on off-resonance noise:

$$L = \frac{1}{N} \sum_{i=1}^N w(f_i) [H_{\text{true}}(f_i) - H_{\text{recon}}(f_i)]^2, \quad w(f) = 1 + (\alpha - 1) G(f) \quad (4.5)$$

where $G(f)$ is the peak-normalised sum of Gaussian bumps centred on each tracked mode’s natural frequency f_m , with width equal to that mode’s half-power bandwidth $\zeta_m f_m$ (floored at 0.5 Hz), read from the Phase 3 baseline model, and $\alpha = 5$ sets how much more heavily the resonance bands are

weighted than the background. If no baseline modal parameters are available, the weighting falls back to a uniform MSE ($w(f) = 1$). Training uses the Adam optimiser (learning rate 10^{-3} , weight decay 10^{-4}) with early stopping on a held-out 20% validation split of the Phase 3 sub-windows (patience 20 epochs, hard ceiling 300 epochs).

4.2.3. Two Health Indicators: HI1 and HI2

Two complementary scalar health indicators are computed per sub-window and aggregated to one value per scan block using the 95th percentile across that block's sub-windows: a statistic robust to occasional noisy sub-windows while remaining more sensitive to genuine damage than the mean and more stable than the maximum.

- *HI1 – resonance-weighted reconstruction error.* The same resonance-weighted MSE used in training (Equation 4.5), evaluated at inference on the current sub-window. A rising HI1 indicates that the observed spectral shape has departed from the pattern the network learned as healthy.
- *HI2 – maximum latent z -score deviation.* For each sub-window, the maximum z -scored deviation, over all 64 latent dimensions, relative to the Phase 3 latent distribution. Because the 64-number bottleneck compresses the spectrum so heavily, a spectral change that is small in full-resolution reconstruction terms can still shift one latent dimension considerably, making HI2 the more sensitive of the two indicators for some forms of subtle early damage, even when HI1 has barely moved.

Both indicators are stored as a two-column array, one row per scan block, in the specimen's HDF5 file (Table 3.5). A healthy run should show both HI1 and HI2 fluctuating near a low, stable floor throughout Phase 3 and the early part of Phase 4, with no sustained upward trend; a successful detection shows one or both rising above their Phase 3 99th-percentile threshold and continuing to rise as damage accumulates. This same 99th-percentile threshold is shared with the live-monitoring system's CAE-based alarm tests (Section 4.6).

Running Layer 0.5 is optional: Layers 1–3 run correctly without it, simply without the two autoencoder-derived damage index formulations D_{CAE} and $D_{\text{CAE,latent}}$ described in Section 4.3.4. It is, however, run for every Group C1 specimen in this thesis, since the comparison between learned and hand-engineered indicators is itself part of the analysis addressing Research Question R2.

4.3. From Many Indicators to One Damage Index

Sections 4.1 and 4.2 each produce a set of candidate damage indicators, computed at every block for every Group C1 specimen: 33 hand-engineered features (20 point-tier, 13 spatial-tier) from Layer 0, computed for the two modes actually tracked with acceptable stability across the Group C1 campaign (Chapter 5, Section 5.2), and, where run, the two learned health indicators HI1 and HI2 from Layer 0.5. This section addresses the two questions that follow once this pooled feature matrix is available. The first is Research Question R2 directly: which of these indicators are actually sensitive to damage, and from what point in specimen life? Three complementary analyses answer this in turn – a correlation analysis quantifying monotonic sensitivity and redundancy, a principal component analysis revealing the intrinsic structure of the feature space, and a SHAP-based supervised ranking with explicit early-life sensitivity. The second question follows directly from the first: having identified which indicators matter and how they relate to one another, how should they be combined into a single, continuous damage index $D(\tau) \in [0, 1]$ that is bounded, physically interpretable, and directly usable as the input to the remaining-life estimation of Section 4.4? All analyses in this section operate on a pooled feature matrix assembled from the eight Group C1 specimens after normalising the time axis to the cycle fraction $\tau = N/N_f$, which removes the natural scatter in total fatigue life before cross-specimen comparisons are made. The two Group C2 specimens are excluded from this pooled analysis for the reasons given in Section 3.5.5.

4.3.1. Damage Sensitivity and Redundancy Analysis

The first component quantifies how strongly each feature tracks damage accumulation and identifies redundant feature pairs that would introduce collinearity into subsequent analyses. For each feature

f and each specimen s , the Spearman rank correlation coefficient ρ_s between the feature time series $f(\tau)$ and the cycle fraction τ is computed:

$$\rho_s(f) = 1 - \frac{6 \sum_i d_i^2}{n(n^2 - 1)} \quad (4.6)$$

where d_i in Equation 4.6 is the difference in ranks between $f(\tau_i)$ and τ_i across the n blocks of that specimen. Spearman correlation is preferred over Pearson here because it is robust to the nonlinear, S-shaped relationship that characterises fatigue indicator evolution – it measures monotonic association rather than linear association, and a feature that increases monotonically with consumed life fraction (even non-linearly) will yield $|\rho_s| \approx 1$ regardless of the precise shape of the trajectory (Conover, 1999). The per-specimen ρ_s values are aggregated as median $\pm$ IQR across the eight Group C1 specimens, revealing which features are consistently sensitive across specimens and which are sensitive only for some (possibly because crack initiation location or surface condition varies).

Mutual information $I(f; \tau)$ is computed alongside ρ_s as a non-parametric complement for features that may exhibit threshold or step-change behaviour rather than monotonic evolution. Features that show high mutual information but low $|\rho_s|$ are flagged for closer inspection, as they may be particularly sensitive in a specific phase of the fatigue life.

Redundancy is assessed separately from sensitivity: the pairwise Pearson correlation matrix between all features is computed on the pooled, normalised feature matrix. Feature pairs with $|r_{ij}| > 0.90$ are classified as redundant; for each redundant pair, the feature with lower median Spearman sensitivity is excluded from the damage index and RUL inputs described later in this chapter. This redundancy pruning prevents ill-conditioned covariance matrices in the anomaly detection model and reduces the risk of overfitting in the Gaussian Process regression of Section 4.4. The excluded features remain in the dataset and are reported in the ranked indicator table for completeness.

4.3.2. Principal Component Analysis

The second component applies Principal Component Analysis (PCA), introduced as a damage-index technique in Section 2.8 of the Literature Review, to the mean-centred, unit-variance normalised pooled feature matrix $\tilde{\mathbf{X}} \in \mathbb{R}^{N_b \times p}$, where N_b is the total number of blocks across all specimens (Jolliffe & Cadima, 2016). Unit-variance normalisation is essential because the features span different physical units and scales; without it, high-variance features such as frequency (in Hz) would dominate the principal directions purely by magnitude.

Two of the resulting principal components (PCs) are of primary interest here. For a well-designed feature set that tracks a single progressive damage process, the first PC (PC1) should capture the dominant direction of feature co-variation across the fatigue life. The loading vector $\mathbf{v}_1$ reveals which features contribute most to this direction, and the fraction of total variance explained, $\lambda_1 / \sum_j \lambda_j$, quantifies how coherently the full feature set tracks a single latent variable (the damage state). Values above 0.80 are interpreted as evidence that a single dominant damage process underlies the observed feature variation; lower values suggest that either multiple damage processes are active or that some features carry substantial noise unrelated to damage (Jolliffe & Cadima, 2016). The PC1 score time series $z_1(\tau) = \tilde{\mathbf{X}}\mathbf{v}_1$ is used as a data-driven composite damage signal in the damage index formulation of Section 4.3.4.

The second PC, by contrast, often captures environmental variability rather than damage: temperature drift, shaker impedance variation, and clamping condition changes are expected confounds for a multi-hour fatigue test. If the PC2 loading vector is dominated by the RMS force and coherence features (which are sensitive to excitation level changes), this confirms that PC2 represents an excitation confound rather than a structural change, and that the PC1 damage signal is robust to this variability.

4.3.3. SHAP-Based Indicator Ranking

The third component provides a supervised, prediction-oriented ranking of indicator importance that directly addresses R2 at the operational level: which features enable the earliest reliable prediction of the remaining fatigue life fraction? A gradient-boosted regression tree ensemble (scikit-learn `GradientBoostingRegressor`, 400 trees, maximum depth 3, learning rate 0.05) is trained to predict the normalised cycle fraction $\tau = N/N_f$ from the feature vector $\mathbf{x}$, using the pooled data from all eight Group C1 specimens as the training set. The gradient-boosted tree is chosen over linear regression because it captures the nonlinear and threshold relationships that characterise individual indicator evolution – many features remain near their baseline values for the first 30–50% of specimen life and then change rapidly; a linear model cannot represent this structure. The model is trained with five-fold cross-validation to assess generalisation and to confirm that the feature importance rankings are not artefacts of a single train-test split; a low training R^2 is expected and unproblematic where only a single specimen or a small pooled set is available, and is logged as an informational note rather than treated as a fitting failure.

SHAP (SHapley Additive exPlanations) values decompose each model prediction into additive per-feature contributions, on the axiomatic grounding established in Section 2.8 of the Literature Review (Lundberg & Lee, 2017). For a prediction at block t for specimen s , the SHAP value $\varphi_j(t, s)$ represents the marginal contribution of feature j to the deviation of that prediction from the mean prediction across the dataset.

Two importance scores are computed from the SHAP values for each feature j :

$$\Phi_j^{\text{global}} = \frac{1}{N_b} \sum_{t,s} |\varphi_j(t, s)| \quad (4.7)$$

$$\Phi_j^{\text{early}} = \frac{1}{|\mathcal{B}_{\text{early}}|} \sum_{(t,s) : \tau < 0.2} |\varphi_j(t, s)| \quad (4.8)$$

where $\mathcal{B}_{\text{early}}$ in Equation 4.8 is the set of blocks with $\tau < 0.2$, relaxed to $\tau < 0.4$ where fewer than 20 blocks would otherwise be available, so that the early-life estimate remains statistically meaningful for smaller specimen populations. Equation 4.7 aggregates the same per-block contributions across the full fatigue life. The global importance Φ_j^{global} ranks features by their average predictive contribution across the full fatigue life. The early-life importance Φ_j^{early} specifically ranks features by their contribution during this early window – the phase in which early detection is most valuable and in which the difference between a useful and a useless indicator is most consequential. A feature with high Φ_j^{early} but low Φ_j^{global} is an early warning indicator that becomes less discriminative later in life; a feature with the reverse profile is a late-stage indicator.

These two aggregate numbers are, in fact, the only SHAP-derived quantities computed by this analysis: a feature's informativeness is summarised as its average contribution across the full fatigue life and its average contribution within the early-life window, rather than as a continuous function of τ . The two together still answer the "when" component of R2 at a coarse resolution: a feature with high Φ_j^{early} relative to Φ_j^{global} is informative early; the reverse profile marks a late-stage indicator. A fully time-resolved $\varphi_j(\tau)$ evolution is not computed by this pipeline stage.

The final output of this analysis is a ranked indicator table with one row per feature, containing: median Spearman ρ_s across specimens, IQR of ρ_s , PC1 loading, global SHAP importance Φ_j^{global} , early-life SHAP importance Φ_j^{early} , redundancy flag, and a composite rank. The composite rank is computed by converting each of ρ_s , Φ_j^{global} , and Φ_j^{early} into a rank position across all features (1st, 2nd, ...), then averaging the three rank positions for each feature, rather than averaging the normalised metric values themselves; rank-averaging is more robust to any one metric having a skewed distribution across features than averaging raw normalised values would be, and ties in the resulting composite rank are possible and are broken using $|\rho_s|$ for sort order only, without changing the displayed rank value. This table is the primary scientific deliverable for Research Question R2 and is formatted for direct inclusion in the Results chapter.

4.3.4. Multiple Damage Index Formulations

The physical motivation for a bounded, monotonically increasing index rooted in the CDM damage variable D was established in Section 2.1; the implementation choices are described here.

Six damage index formulations are computed in parallel: four hand-engineered formulations derived directly from the feature set and sensitivity analysis above, and two further formulations derived from the Layer 0.5 autoencoder health indicators (Section 4.2). Computing multiple formulations is not redundant: each formulation encodes a different aspect of the damage state – statistical distance from baseline, stiffness loss, data-driven feature consensus, importance-weighted sensitivity, and, for the two autoencoder-derived formulations, an entirely learned representation of healthy-state deviation that makes no prior assumption about which spectral features matter – and comparing them across the eight Group C1 specimens reveals which representation is most monotonic, most consistent across specimens, and most predictive of the failure point. This comparison is itself a scientific contribution: it provides empirical guidance for future monitoring system designers on which damage representation to prioritise, and, for the autoencoder-based pair specifically, on whether a learned representation is competitive with hand-engineered features in this experimental setting.

D_{T^2} – Statistical distance formulation. The Hotelling T^2 score at each block (Section 4.1.7) is normalised to $[0, 1]$ by dividing by the T^2 value at the failure block:

$$D_{T^2}(\tau) = \frac{T^2(\tau)}{T^2(\tau_f)} \quad (4.9)$$

Equation 4.9 integrates information from the full feature vector simultaneously and is sensitive to any deviation from the healthy baseline, regardless of its physical origin. Its limitation is that the normalisation requires knowledge of the final T^2 value, making it retrospective rather than online; it cannot be computed until the test is complete. It is therefore used as a benchmark formulation for post-hoc comparison rather than as a live prognostic index.

D_{freq} – Frequency-based formulation. The relative natural frequency shift at the primary mode, sign-flipped to be positive-increasing:

$$D_{\text{freq}}(\tau) = \frac{f_{n,0} - f_n(\tau)}{f_{n,0} - f_{n,\min}} = \frac{-\Delta f_n(\tau)}{|\Delta f_{n,\max}|} \quad (4.10)$$

where $f_{n,0}$ in Equation 4.10 is the Phase 3 baseline frequency and $f_{n,\min}$ is the minimum frequency observed across the specimen's life. This formulation is directly linked to relative stiffness loss through the CDM relationship $\Delta\omega_n/\omega_{n,0} \approx -\frac{\alpha}{2}D$ (Equation 2.5), making it the most physically interpretable index. It is the only formulation that can be computed in real time without retrospective normalisation if an estimated $f_{n,\min}$ is available, and it is monotonic for the majority of specimens. Its limitation is potential early saturation if the frequency shift plateaus before fracture, or non-monotonic behaviour in the early initiation phase due to the initial material hardening effect reported for aluminium specimens (Lorenzino & Navarro, 2015).

D_{PC1} – PCA-based formulation. The projection of the feature vector onto the first principal component (established in Section 4.3.2), normalised to $[0, 1]$:

$$D_{\text{PC1}}(\tau) = \frac{z_1(\tau) - z_{1,\min}}{z_{1,\max} - z_{1,\min}} \quad \text{where} \quad z_1(\tau) = \tilde{\mathbf{x}}(\tau)^\top \mathbf{v}_1 \quad (4.11)$$

and $\mathbf{v}_1$ in Equation 4.11 is the PC1 loading vector from the PCA above. This formulation captures the dominant direction of change across the full feature space and is the most robust to individual noisy features: because PC1 aggregates evidence from all indicators, noise in one channel is diluted by the consistent signal in others. Its limitation is that PC1 is computed post-hoc and is not directly interpretable in terms of a single physical quantity. In practice, if the feature set is well-designed and

the experiment is dominated by a single progressive damage process, D_{PC1} and D_{freq} should closely agree – their divergence is informative about the extent to which frequency shift alone captures the full damage state.

D_{comp} – Composite formulation. A weighted sum of the top- k non-redundant features identified by the SHAP ranking above, normalised to $[0, 1]$:

$$D_{comp}(\tau) = \text{normalise} \left(\sum_{j \in \mathcal{S}_k} w_j \tilde{x}_j(\tau) \right) \quad (4.12)$$

where $\mathcal{S}_k$ in Equation 4.12 is the set of the top- k non-redundant features by global SHAP importance, w_j are the SHAP-derived weights, and normalisation maps the result to $[0, 1]$ over the specimen's life. The composite formulation is designed to maximise damage sensitivity by concentrating weight on the most informative features as identified by the SHAP analysis; it requires the full sensitivity analysis above to be completed before the weights can be computed and is therefore also retrospective in this study. With sufficient training data from prior specimens, the SHAP weights could in principle be transferred to new in-progress tests, making D_{comp} an online index for future campaigns.

D_{CAE} and $D_{CAE,latent}$ – Autoencoder-based formulations. The two Layer 0.5 health indicators, HI1 (resonance-weighted reconstruction error) and HI2 (maximum latent z -score deviation), are each forward-filled from scan cadence to full block cadence and min-max normalised to $[0, 1]$ over the specimen's life:

$$D_{CAE}(\tau) = \frac{HI1(\tau) - \min(HI1)}{\max(HI1) - \min(HI1)} \quad (4.13)$$

and analogously for $D_{CAE,latent}(\tau)$ from HI2, Equation 4.13 rescaling each indicator onto $[0, 1]$ over the specimen's observed range. Unlike the four formulations above, neither is derived from a hand-chosen feature or a linear/Gaussian statistical model: both are derived from a convolutional autoencoder trained exclusively on Phase 3 (healthy) log-magnitude FRF sub-windows, so that any departure from the learned healthy manifold – whether or not it corresponds to a feature the hand-engineered pipeline tracks – registers as a rising health indicator (Section 4.2). D_{CAE} and $D_{CAE,latent}$ participate in the same logistic S-curve fitting, monotonicity analysis, and per-specimen prognostic horizon calculation (Section 4.4) as the four formulations above, and are included as candidate formulations for the S-N lead-time overlay (Section 4.5). They are also included, alongside D_{T^2} , D_{freq} , D_{PC1} , and D_{comp} , in the population-level Gaussian Process model of Section 4.4.3, so that the GP-based prognostic horizon analysis is available for all six formulations on an equal footing; an earlier implementation defect that temporarily restricted the GP to four of the six formulations is noted and corrected in Section 4.4.4. Running Layer 0.5 is optional and computationally heavier than the rest of Layer 0 (it requires a GPU-capable PyTorch installation for practical training times); where it was not run for a given specimen, D_{CAE} and $D_{CAE,latent}$ are reported as unavailable for that specimen rather than substituted with a default value.

The six formulations are evaluated against three criteria: *monotonicity* (is $D(\tau)$ consistently non-decreasing across all blocks?), *sensitivity onset* (at what cycle fraction τ_{onset} does $D(\tau)$ first exceed the Phase 3 variance band by more than 3σ ?), and *inter-specimen consistency* (how similar are the DI curves across the eight Group C1 specimens when plotted on the τ axis?). These comparisons are reported in the Results chapter.

4.3.5. Logistic Trend Fitting and Physical Interpretation

For each specimen and each formulation, the logistic (S-curve) model is fitted to the complete DI time series using nonlinear least squares:

$$\hat{D}(\tau; \beta, \tau_{mid}) = \frac{1}{1 + \exp(-\beta(\tau - \tau_{mid}))} \quad (4.14)$$

where $\beta > 0$ controls the steepness of the damage acceleration and $\tau_{\text{mid}} \in (0, 1)$ is the inflection point – the cycle fraction at which damage is at 50% of its failure-stage value. The logistic model is chosen because it is physically consistent with Paris-law crack growth dynamics (which produce an S-shaped damage trajectory, as established in Section 2.1), it has a natural upper asymptote at $D = 1$ that prevents unphysical extrapolations, and it is parameterised by two quantities that have direct physical interpretations.

The fitted parameters $\hat{\beta}$ and $\hat{\tau}_{\text{mid}}$ are compared across specimens and DI formulations. Consistency in $\hat{\tau}_{\text{mid}}$ across specimens indicates that the damage process has a characteristic timing relative to total life; scatter in $\hat{\tau}_{\text{mid}}$ reflects inter-specimen variability in crack initiation timing. The steepness parameter $\hat{\beta}$ characterises the damage acceleration rate: larger values indicate a more abrupt transition from slow initiation to rapid propagation. These fitted parameters enter the RUL prognosis of Section 4.4 as the starting point for the online extrapolation.

4.3.6. Inter-Specimen Comparison

All eight Group C1 DI curves for each formulation are plotted on the normalised cycle fraction axis $\tau = N/N_f$, which removes the natural life scatter between nominally identical specimens before the shapes are compared. This normalisation is essential: without it, a specimen with $N_f = 120,000$ cycles would appear to damage more slowly than one with $N_f = 180,000$ cycles, even if both follow identical $D(\tau)$ trajectories. On the τ axis, the band of DI curves gives a direct measure of the uncertainty in damage index evolution across the population – an uncertainty that propagates directly into the confidence intervals on remaining useful life described in Section 4.4. The two Group C2 specimens are reported separately (Section 3.5.5) rather than included in this band, since their interrupted Phase 4 protocol makes their $D(\tau)$ trajectories not directly comparable to the uninterrupted Group C1 record.

Expected sources of inter-specimen variability include surface finish at the notch root (which controls crack initiation timing), minor clamping alignment differences (which affect the stress distribution near the fixture), and natural statistical scatter in fatigue crack nucleation and propagation for metallic specimens (for which N_f typically varies by a factor of two to three even under nominally identical conditions (Schijve, 2008)). The spread of the DI band across specimens is therefore both an expected and a physically informative result.

4.3.7. Comparison with Synthetic Data Toolkit Predictions

As described in Section 3.1, the Synthetic Data Toolkit produces a predicted $D(\tau)$ trajectory from the CDM–Basquin fracture mechanics model for the final specimen geometry and material. After the fatigue campaign is complete, the experimental D_{freq} curves are overlaid on the toolkit prediction for the same kind of post-hoc, non-validating comparison described in Section 3.1.4; here the comparison is specifically against the simplified Paris-law growth model underlying the toolkit's damage kinematics, rather than against the toolkit's boundary-condition or geometry assumptions.

Where the experimental curves agree with the synthetic prediction, particularly in the shape of the S-curve and the approximate location of τ_{mid} , this provides confidence that the CDM framework is an adequate description of the dominant damage kinematics. Where they diverge, the divergence is scientifically informative: it identifies the physical mechanisms not captured by the simplified model and motivates the discussion in Chapter 6. Crucially, the experimental data are never fitted to the synthetic model; the two evolve independently and are compared post-hoc, preserving the integrity of both.

4.4. Remaining Useful Life Estimation (Layer 3)

Layer 3 addresses Research Question R3 directly: given the damage index history accumulated up to the current fatigue block, how many cycles remain until specimen failure, and how uncertain is that estimate? Two methods are implemented in sequence, with complementary strengths. Method 1 (parametric trend extrapolation) operates on the data from the specimen currently under test and is applicable from the first block onward. It requires no prior specimen data but its uncertainty is high early in life when only a small portion of the DI curve has been observed. Method 2 (Gaussian Process regression) uses the run-to-failure records from all previously completed specimens as training data, providing a calibrated, data-driven uncertainty estimate that is available for any new specimen tested

after the campaign is complete. Both methods share the same failure threshold, prognostic horizon definition, and evaluation metrics, enabling direct comparison.

A key principle underlying both methods is that RUL estimation is inherently uncertain, and that the scientifically meaningful question is not whether a single prediction is accurate but from what point in specimen life a useful prediction becomes available. The *prognostic horizon* τ_h captures this operationally relevant quantity and is the primary performance metric for Layer 3.

4.4.1. Failure Threshold

The failure threshold $D_f = 0.9$ is adopted across all DI formulations and both RUL methods. Extrapolation to D_f rather than to $D = 1$ is motivated by the observed behaviour of the logistic DI model: the DI approaches unity asymptotically and physical fracture typically occurs while D is still rising steeply, before the asymptote is reached. Setting $D_f = 0.9$ corresponds approximately to the point of unstable crack propagation where the remaining life is negligible relative to the total life, consistent with the failure criterion used in comparable studies (Eltouny et al., 2023). The RUL at any block is then:

$$\text{RUL}(\tau) = \tau_f - \tau \quad (4.15)$$

where τ_f in Equation 4.15 is the predicted cycle fraction at which $D = D_f$ and τ is the current cycle fraction. All RUL values are reported in units of normalised cycle fraction; conversion to absolute cycles requires multiplication by the estimated total life $\hat{N}_f$, which is itself updated as the logistic fit is refined.

4.4.2. Method 1: Parametric Trend Extrapolation

Online logistic model fitting

At each new Phase 4 block, the logistic model (Equation 4.14) is refitted by nonlinear least squares to all DI observations accumulated up to and including the current block:

$$\{\hat{\beta}(\tau), \hat{\tau}_{\text{mid}}(\tau)\} = \arg \min_{\beta, \tau_{\text{mid}}} \sum_{k=1}^{n(\tau)} [D(k) - \hat{D}(k; \beta, \tau_{\text{mid}})]^2 \quad (4.16)$$

where $n(\tau)$ in Equation 4.16 is the number of blocks observed up to the current cycle fraction. The model is initialised with $\beta_0 = 10$ and $\tau_{\text{mid},0} = 0.7$ – values that place the initial prediction at a conservative late-life position, avoiding premature alarm. The refitted parameters are used to extrapolate the DI trajectory forward to $D_f = 0.9$, giving $\hat{\tau}_f$ and hence $\text{RUL}(\tau)$.

Uncertainty quantification

The 95% confidence interval on $\hat{\tau}_f$ is propagated from the parameter covariance matrix $\mathbf{C}(\tau)$ of the nonlinear least-squares fit, estimated from the Jacobian at the solution. This linearised (delta-method) propagation is used in preference to a bootstrap or Monte Carlo alternative because it must be re-computed at every new block as the fit is refreshed online: the Jacobian is already available as a by-product of the nonlinear least-squares solver, so the covariance follows at negligible extra cost, whereas a resampling-based interval would need many refits per block and is not practical at this update cadence.

$$\text{Var}(\hat{\tau}_f) \approx \nabla_{\theta} \tau_f \Big|_{\hat{\theta}}^{\top} \mathbf{C}(\tau) \nabla_{\theta} \tau_f \Big|_{\hat{\theta}} \quad (4.17)$$

where $\theta = [\beta, \tau_{\text{mid}}]^{\top}$ and the gradient in Equation 4.17 is evaluated numerically. The CI width narrows monotonically as more blocks are observed: early in life, τ_{mid} is poorly constrained because only the lower tail of the S-curve has been observed, producing wide intervals; as the inflection point is passed, the fit stabilises rapidly.

Early-life fallback: power-law model

When fewer than five Phase 4 blocks have been accumulated, or when the logistic fit fails to converge (typically because $D(\tau)$ has not yet reached 0.5 and the inflection point is unconstrained), a power-law model is used as a fallback. A power law is chosen over, say, a straight-line extrapolation because early-life damage accumulation is itself the shallow lower tail of the same S-shaped process (Section 4.3.4): a power law with $b < 1$ reproduces this decelerating-then-accelerating character locally without requiring the inflection point that the full logistic model needs to be identifiable, while still collapsing to a sensible, bounded estimate rather than the unbounded growth a linear fit would produce if extrapolated far beyond the observed data:

$$D(\tau) = a \tau^b \quad (4.18)$$

with the parameters a and b of Equation 4.18 estimated by linear regression in log-log space. The power-law model is more stable than the logistic for early-life extrapolation because it does not require a visible inflection point; however, it has no natural upper asymptote and can produce unphysical extrapolations at $\tau \rightarrow 1$. It is therefore used exclusively as a transient early-life bridge and replaced by the logistic model as soon as the fit converges. The fallback status of any prediction is stored as a flag in the HDF5 output so that early-life power-law predictions can be distinguished from logistic predictions in the Results analysis.

4.4.3. Method 2: Gaussian Process Regression

Motivation

The parametric extrapolation of Method 1 relies solely on the DI history of the specimen currently under test. It carries no information about how similar specimens have behaved in the past, and its confidence intervals reflect only the parameter uncertainty of the logistic fit rather than the true population variability. Gaussian Process (GP) regression, introduced in Section 2.9 of the Literature Review, addresses both limitations: by training on the run-to-failure records of the other seven completed Group C1 specimens, it incorporates population-level variability into the predictive uncertainty in a principled way (Rasmussen & Williams, 2006).

GP input feature vector

At prediction time τ , the GP receives a fixed-length input vector $\mathbf{x}_{\text{GP}}(\tau) \in \mathbb{R}^6$ constructed from the current damage state and its recent history:

$$\mathbf{x}_{\text{GP}}(\tau) = [D(\tau), \dot{D}(\tau), \tau, \varphi_{j_1}(\tau), \varphi_{j_2}(\tau), \varphi_{j_3}(\tau)]^\top \quad (4.19)$$

where $D(\tau)$ in Equation 4.19 is the composite damage index at the current block, $\dot{D}(\tau) = dD/d\tau$ is its finite-difference derivative (the current rate of damage accumulation), τ is the current cycle fraction, and $\varphi_{j_1}, \varphi_{j_2}, \varphi_{j_3}$ are the current values of the top-3 non-redundant features by global SHAP importance from the Layer 1 ranking. Including the damage rate $\dot{D}$ captures whether the DI is still in the initiation plateau or has entered the acceleration phase – information that the current D value alone cannot distinguish. Including the top SHAP features allows the GP to leverage the raw indicator signals alongside the composite DI, improving predictions in cases where D_{comp} has not yet fully separated from the baseline variance band.

Kernel and hyperparameter optimisation

A Matérn 5/2 kernel is used:

$$k(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \left(1 + \frac{\sqrt{5}r}{\ell} + \frac{5r^2}{3\ell^2} \right) \exp\left(-\frac{\sqrt{5}r}{\ell}\right) \quad (4.20)$$

where $r = \|\mathbf{x} - \mathbf{x}'\|_2$ in Equation 4.20, σ_f^2 is the signal variance, and ℓ is the lengthscale. The Matérn 5/2 kernel assumes the target function is twice differentiable, which is appropriate for smooth damage

progression curves and avoids the over-smoothing tendency of the squared-exponential kernel (Rasmussen & Williams, 2006). The kernel hyperparameters σ_f^2 and ℓ are optimised by maximising the log marginal likelihood on the training data. An independent Gaussian noise term σ_n^2 is added to the kernel diagonal to account for measurement noise and inter-specimen variability not captured by the feature vector.

Leave-one-out cross-validation

The GP model is evaluated using leave-one-out cross-validation (LOO-CV) across the eight Group C1 specimens. For each specimen $s \in \{1, \dots, 8\}$, the GP is trained on the remaining seven completed specimens and used to predict the held-out specimen's RUL at three checkpoints: $\tau = 0.2$, $\tau = 0.4$, and $\tau = 0.6$. At each checkpoint, the prediction is made using only the feature vector available up to that cycle fraction, mirroring the online monitoring scenario. The root-mean-square prediction error (RMSPE) across specimens is computed at each checkpoint:

$$\text{RMSPE}(\tau_c) = \sqrt{\frac{1}{8} \sum_{s=1}^8 [\widehat{\text{RUL}}_s(\tau_c) - \text{RUL}_s^{\text{true}}(\tau_c)]^2} \quad (4.21)$$

where $\text{RUL}_s^{\text{true}}(\tau_c) = 1 - \tau_c$ in Equation 4.21 is the true remaining cycle fraction of specimen s at checkpoint τ_c . The LOO-CV RMSPE at $\tau_c = 0.2$ is the primary early-detection performance metric; RMSPE monotonically decreasing from $\tau = 0.2$ to $\tau = 0.6$ confirms that the GP improves as more damage history is available.

4.4.4. Prognostic Horizon and Evaluation Framework

The prognostic horizon τ_h is defined as the earliest cycle fraction at which the 95% confidence interval on the RUL estimate spans less than 30% of the total expected life (i.e. a half-width of ± 0.15 on the normalised τ axis):

$$\tau_h = \min\{\tau \mid \Delta_{\text{CI}}(\tau) < 0.30\} \quad (4.22)$$

where $\Delta_{\text{CI}}(\tau)$ in Equation 4.22 is the full width of the 95% CI at the current block. The threshold of 0.30 is chosen to correspond to a practically useful level of precision: a prediction that narrows the expected remaining life to within $\pm 15\%$ of total life is actionable for maintenance scheduling, whereas a wider interval is not, consistent with the prognostic horizon concept formalised in the PHM literature (Saxena et al., 2008). The prognostic horizon is computed separately for Method 1 and Method 2 and for each of the six DI formulations, giving a matrix of τ_h values that directly quantifies the improvement offered by the GP approach over simple trend extrapolation, and by the composite DI over the single-indicator D_{freq} . The Method 2 Gaussian Process is trained on all six DI formulations, including D_{CAE} and $D_{\text{CAE,latent}}$. An earlier version of the implementation restricted it to four, and all results reported in Chapter 5 were regenerated after that restriction was removed. If Layer 0.5 was not run for a given specimen, D_{CAE} and $D_{\text{CAE,latent}}$ are returned as all-zero arrays with a logged warning for that specimen, and their S-curve fits will not converge; this did not occur for any Group C1 specimen in this campaign, since Layer 0.5 was run for all eight.

Two quantities constitute the reported answer to R3: the prognostic horizon τ_h , the earliest cycle fraction at which the remaining-life estimate becomes sufficiently confident to be actionable, and the prediction error at that horizon. Both are reported as median and interquartile range across the eight Group C1 specimens rather than for any single specimen, so that the result characterises the population rather than the idiosyncrasies of one test. Reporting them in this form does not presuppose that a horizon is reached: a formulation for which the confidence criterion is never satisfied yields no τ_h , and that outcome is itself the reported result for that formulation. Both Method 1 and Method 2 prognostic horizons are evaluated for all six DI formulations (Section 4.3.4).

4.5. The S-N Curve Module

The S-N module was developed after the core Layer 0–3 pipeline, specifically to give Research Question R4 a dedicated quantitative analysis rather than relying solely on qualitative synthesis across the other layers. It converts the fixed-frequency fatigue record already collected for every specimen into a classical Basquin stress-life characterisation, and overlays each specimen’s Layer 3 prognostic horizon directly onto that S-N plot, making visible, on the same axis a fatigue engineer already uses for design, how much of the conventional S-N scatter band the monitoring pipeline’s early-warning capability buys back. The overlay is constructed by converting each specimen’s prognostic horizon τ_h from the Layer 3 RUL report (Section 4.4.4) into an absolute cycle count $\tau_h \times N_f$, plotting it as a lead-time marker at that specimen’s stress level, and drawing a horizontal arrow from the marker to the specimen’s actual failure point (N_f, S_a) , so that the arrow length reads directly as prognostic lead time in cycles. A specimen and formulation combination that never reached τ_h carries no marker, which is itself informative. The resulting plot and its interpretation are reported in Section 5.6.4 of Chapter 5.

The eight Group C1 specimens were tested at three deliberately chosen target force levels, with two to three specimens per level, a design established during the Group B load configuration study (Section 3.4.3) specifically so that the resulting campaign would support a Basquin S-N fit alongside its primary role as the modal-parameter monitoring dataset. The five Group B specimens themselves, tested under a different protocol during that tuning phase rather than the final campaign design, are excluded from the fit described below and retained instead as contextual information for the Discussion chapter.

Each specimen is driven by a harmonic force $F(t) = F_m + F_a \sin(\omega t)$. The amplitude and mean are extracted directly from the raw force signal recorded for the first Phase 4 block, chosen because it is always retained regardless of the raw-signal retention interval used elsewhere in the pipeline, and because the constant-amplitude harmonic protocol means it is representative of the steady-state loading condition for the specimen’s entire fatigue life:

$$F_a = \frac{\max(F) - \min(F)}{2}, \quad F_m = \text{mean}(F), \quad R = \frac{F_m - F_a}{F_m + F_a} \quad (4.23)$$

A scale factor k converts the force amplitude of Equation 4.23 to a nominal stress amplitude, $S_a = kF_a$. Rather than leaving k at its pipeline default of 1.0 (force units, pending calibration), k was calibrated for this campaign on a per-specimen basis: a static finite element analysis was run in FEMAP for each specimen, applying that specimen’s own measured force amplitude as the boundary load and extracting the resulting stress at the notch root, giving $k_{\text{specimen}} = \sigma_{\text{FEM}}/F_{\text{FEM}}$ individually rather than a single value assumed to hold across the population. This per-specimen approach absorbs any real differences in effective stiffness, notch stress concentration, or clamping condition between nominally identical specimens directly into the calibration, rather than allowing such differences to be folded into the fitted Basquin slope as if they were genuine scatter in the material’s fatigue response. The calibrated k values themselves, together with the corresponding FEMAP stress field results, are reported in Chapter 5 alongside the fitted S-N curve they produce.

Whereas the Synthetic Data Toolkit (Section 3.1) uses the Basquin law of Section 2.1 predictively, to drive the simulated damage increments, here the same functional form is instead fitted empirically to the observed (S_a, N_f) pairs from the completed campaign. The model is fitted in log-log space by ordinary least squares across every specimen with valid force-amplitude data and $N_f > 0$:

$$\log_{10}(N) = \log_{10}(C) - b \log_{10}(S_a) \quad (4.24)$$

A note on notation is needed here, since the same symbol carries two conventions in the literature. The strain-life form quoted in Section 2.1.1, $N_f = (\sigma_a/\sigma'_f)^{1/b}$, places the exponent in the denominator, so its b takes small negative values, conventionally -0.05 to -0.15 for metallic alloys. The stress-life form fitted here places the exponent directly on S_a , so the two are related by $b_{\text{here}} = -1/b_{\text{Section 2.1.1}}$ and the literature range corresponds to a b of approximately 7 to 20 in the present convention. All values of b reported in Chapter 5 use the convention of Equation 4.24.

Equation 4.24 is accompanied by a 95% prediction band, bounding individual future observations, not the mean, computed analytically from the standard OLS formula. A meaningful fit requires specimens tested at more than one distinct stress level; distinct levels are identified by a simple gap rule (sorted S_a values are grouped together unless consecutive values differ by more than 5% of the maximum observed S_a), and if fewer than three distinct levels are detected, the fit is still computed and plotted but is flagged as preliminary. The three-level design described above is intended to satisfy this requirement comfortably, though with only two to three repeats per level the fitted slope b still carries a wider confidence interval than a dedicated S-N campaign with more repeats per level would provide (Section 7.4.1 of Chapter 7).

4.6. Live Monitoring

Every analysis described so far in this chapter is a batch process: a complete set of blocks is loaded from a specimen's HDF5 file and processed in one pass. The live-monitoring system addresses a narrower and operationally distinct question – whether, within seconds of a fatigue block completing on the rig, the specimen has just crossed a damage or fracture threshold – without waiting for the full Layer 0–3 pipeline to be re-run. It is a convenience layer over the same physical models and statistical thresholds used throughout the rest of the pipeline rather than an independent detection method, and using it has no effect on the batch Layer 0–3 results: it is optional, and was used during this campaign both as a general real-time monitoring aid and, specifically for the two Group C2 specimens, as the mechanism that determined when Phase 4 loading should be interrupted for microscope re-inspection (Section 3.5.5).

4.6.1. Polling and Block Processing

Once at least one specimen has completed Layer 0 in the working session – so that the Hotelling baseline model and its detection thresholds exist to compare against – the live monitor can be pointed at a specimen's data folder while a test is actively running. It polls that folder on a 15-second interval; when a new, complete block folder is detected, a reduced version of the Layer 0 (and, if available for that specimen, Layer 0.5) feature extraction is run on that single block alone and scored against the pre-fitted baseline models, rather than re-running the full batch pipeline over the entire history. If Layer 0.5 has not yet been run for the monitored specimen, the CAE-based tests below are simply skipped for that session rather than treated as an error: a specimen can be live-monitored using the statistical tests alone before Layer 0.5 has ever been run for it.

4.6.2. Two Alarm Classes

Two alarm levels are distinguished, reflecting two different physical events that a live monitor should respond to differently.

- **Damage alarm.** Fires when any one of up to four independent statistical tests exceeds its threshold for the newly processed block: Hotelling T^2 (the same flagged/not-flagged decision as the batch Layer 0 detector, Section 4.1.7, applied to the one new block); HDI exceeding its Phase 3 baseline mean by more than five standard deviations; and, where Layer 0.5 has been run for the monitored specimen, HI1 or HI2 exceeding its stored Phase 3 99th-percentile threshold (Section 4.2.3). These four tests are not combined into a single score; any one is sufficient to raise the alarm, and the alarm names which specific test fired. This is a deliberate choice for a live monitor used by a researcher acting in real time: knowing immediately which evidence triggered an alarm is more useful than waiting for a combined score that obscures its source. A single, isolated damage alarm is not necessarily urgent, since the same short-scan and environmental noise sources discussed elsewhere in this pipeline can trigger any one test in isolation; a genuine damage onset is expected to produce sustained or escalating alarms across several subsequent blocks, ideally across more than one of the four tests.
- **Fracture alarm.** Fires when the same within-block, Hilbert-envelope amplitude-collapse detector used by the batch Phase 4 processing finds a sustained anomaly in the new block's raw velocity signal consistent with the specimen having just fractured – the same physical event and detection logic as the batch fracture detector, evaluated immediately rather than after the full test has been

loaded. Specifically, the detector requires a fractional amplitude collapse of at least 0.2 relative to the block's expected level, sustained across at least three consecutive analysis windows for a minimum duration of 0.5 s, with the first 5% of the block excluded from consideration to avoid reacting to startup transients. A fracture alarm is treated as actionable immediately, in contrast to a damage alarm: the detector is specifically designed, via this minimum sustained-duration requirement, to avoid firing on short transients or cable artefacts, so a fracture alarm should not be dismissed as noise.

For the two Group C2 specimens, the damage alarm, not the fracture alarm, is the event that determines the Stage 2/Stage 3 transition in the fractography validation protocol (Section 3.5.5): loading is interrupted at the first damage alarm, and the specimen is returned to the microscope before any further fatigue cycling resumes. The live monitor's outputs are designed to agree with what the batch Layer 0 and Layer 0.5 analyses would report for the same block if the full pipeline were re-run after the test, providing a useful internal consistency check between the real-time and batch analysis paths.

5

Results

This chapter reports the experimental outcomes of the fatigue monitoring campaign described in Chapter 3 and analysed according to the methodology of Chapter 4, organised around the four research questions set out in Chapter 1. Results are presented without interpretation: comparison with the literature, discussion of physical mechanisms, and assessment of the broader implications of these findings are deferred to Chapter 6. Where a result requires a definition or threshold to be stated for the figure or table to be read correctly (for example, the cycle fraction $\tau = N/N_f$, or a detection threshold established in Chapter 3), that definition is restated briefly for clarity, but no claim is made here about what the result means.

Section 5.1 establishes, in chronological order, the pre-campaign constants that later sections rely on, the reason Group B is excluded from the population results, and the Group C specimen population and its completeness. Section 5.2 reports the modal parameter evolution records (R1). Section 5.3 reports the indicator sensitivity ranking (R2). Section 5.4 reports the damage index trajectories constructed in Layer 2, which bridge R1–R3 rather than answering a single research question on their own. Section 5.5 reports the remaining-useful-life and prognostic horizon analysis (R3). Section 5.6 reports the S-N characterisation and resonance-testing context (R4). Section 5.7 reports the independent optical fractography validation conducted on the two Group C2 specimens. Table 5.1 summarises this mapping.

Table 5.1: Mapping of chapter sections onto the research questions they address.

Section	Content	RQ
5.1 Specimen Overview and Data Completeness	Pre-campaign constants, Group B exclusion, Group C population and completeness	—
5.2 Modal Parameter Evolution	How the vibration response evolves as damage accumulates	R1
5.3 Indicator Sensitivity Ranking	Which indicators are most sensitive, hand-engineered and autoencoder-based	R2
5.4 Damage Index Trajectories	How the ranked indicators combine into a single $D(\tau)$	R1–R3 (bridging)
5.5 Remaining Useful Life Estimation	How much life remains, and from what point that estimate is reliable	R3
5.6 S-N Characterisation and Resonance Testing Context	How this campaign’s fatigue life compares to classical stress-life testing	R4
5.7 Independent Validation via Optical Fractography	Whether detection corresponds to a physically observed event	Validation

5.1. Specimen Overview and Data Completeness

The pre-campaign sensor characterisation battery (Section 3.4.1) establishes three instrument-level constants referenced throughout this chapter. The LDV velocity noise floor (Test T1, Figure A.1 of Chapter 3) rises from below 10^{-14} (m/s)²/Hz at low frequency to a broadband plateau of 10^{-13} – 10^{-12} (m/s)²/Hz through the 10–1000 Hz range spanning the first three bending modes, integrating to approximately 1.0×10^{-5} m/s of cumulative RMS noise over that band; this is the reference floor against which the earliest detected indicator onsets in Sections 5.3 and 5.4 are judged. Test T2 (Figure A.2 of Chapter 3) confirms that the 30 Hz Phase 3/4 drive frequency sits well within the range over which both the LDV and the reference accelerometer perform adequately, bounding rather than directly measuring detection performance at the operating frequency. Test S4 established the onset of measurable nonlinearity at a drive level of approximately 0.3 V (Figure 3.14 of Chapter 3), confirming that the broadband scanning excitation used throughout Phase 3 and Phase 4 remains within the linear regime assumed by every FRF-based feature reported in Sections 5.2–5.4; the harmonic loading stage itself is, by design, mildly nonlinear (Section 3.4), a distinction returned to when the harmonic-stage features are discussed in Chapter 6.

The five Group B specimens used for the load configuration study (Section 3.4.3) are excluded from the population reported in this chapter. Unlike the Group C campaign, Group B was not run under one fixed protocol: rig stiffness, sample placement, and excitation level were each varied iteratively across, and in some cases within, individual specimens as the final configuration was converged upon, with the notch geometry itself changing partway through the study (Section 3.2.2). Group B's fatigue lives are therefore not a population in the sense that Group C1 is, and are not reported here or compared against the Group C1 N_f range. Their contribution to this thesis was methodological, informing the final protocol, rather than result-bearing.

Of the ten Group C specimens fabricated to the final Al 6061-T6 geometry (Section 3.2.3), eight (Group C1) were run to complete fracture under the uninterrupted Phase 4 protocol and constitute the population-level dataset used throughout Sections 5.2–5.6. The remaining two (Group C2) were run under the interrupted protocol and are reported separately in Section 5.7. Table 5.2 summarises, for every Group C specimen, the total Phase 4 (fatigue) cycles to fracture – excluding the five-block, sub-endurance-limit Phase 3 baseline common to every specimen – or to interruption and subsequent final fracture for Group C2, and which sections of this chapter draw on its data. Specimen identifiers map directly onto the sample numbers recorded during the campaign (Sample $N \rightarrow C1-0N$).

Table 5.2: Overview of the ten Group C specimens: Phase-4 (fatigue) cycles to fracture and data availability across this chapter. N_f excludes the five-block Phase 3 baseline common to every specimen and reflects the pipeline's own precise within-block fracture detection (Section 5.6) rather than a block-count estimate. Layer 0.5 was run for all eight Group C1 specimens.

Specimen	Group	N_f (cycles)	Sections using this specimen
C1-01	C1	6,930	5.2–5.6
C1-02	C1	15,000	5.2–5.6
C1-03	C1	11,880	5.2–5.6
C1-04	C1	59,250	5.2–5.6
C1-05	C1	112,500	5.2–5.6
C1-06	C1	24,000	5.2–5.6
C1-07	C1	35,250	5.2–5.6
C1-08	C1	55,860	5.2–5.6
C2-01	C2	$N_{\text{interrupt}} \approx 7,500$; $N_f \approx 7,800$ (final)	5.7 only
C2-02	C2	$N_{\text{interrupt}} = 7,500$; $N_f = 8,200$ (final)	5.7 only

Layer 0.5 was run for all eight Group C1 specimens, so the $D_{\text{CAE}}/D_{\text{CAE,latent}}$ formulations and the H11/H12 comparison in Sections 5.3–5.4 are based on the same full population as the hand-engineered

indicators, with no $n < 8$ caveat required.

For the two Group C2 specimens, the interrupted protocol produced markedly different outcomes: C2-01 was interrupted at $N_{\text{interrupt}} \approx 7,500$ cycles into Phase 4 and, after microscope re-inspection, fractured almost immediately upon resumption (within approximately 10 seconds, or roughly 300 cycles at the 30 Hz drive frequency, into the first resumed block), giving a final $N_f \approx 7,800$ cycles and essentially no residual life at the point of interruption. C2-02 was interrupted at a similar damage progression, at $N_{\text{interrupt}} = 7,500$ cycles into Phase 4, and fractured after a further 700 cycles of resumed loading, at $N_f = 8,200$ cycles.

5.2. Modal Parameter Evolution (R1)

Research Question R1 asks: *how does the vibration response evolve as fatigue damage accumulates under resonant vibration loading?* Before reporting that evolution, one data-quality result must be stated up front, since it constrains everything that follows: across the eight Group C1 specimens, only two modes could be identified with acceptable stability by the modal parameter estimation routine, not the three anticipated in the experimental design (Chapters 3–4). A putative third mode was visible in places but could not be tracked consistently enough to be included. Even the two retained modes show block-to-block pole fluctuation beyond what a stable identification would be expected to show. All results in this section and in Sections 5.3–5.6 are therefore reported for Mode 0 and Mode 1 only, using the specimen labels S01–S08 as recorded by the analysis pipeline, which correspond directly to C1-01–C1-08 (Table 5.2). Whether Mode 0 and Mode 1 correspond to the same physical bending mode across all eight specimens could not be confirmed independently at the achieved measurement resolution; Section 5.2.6 reports MAC evidence indicating that, at least in some blocks, they do not. No cause for the fluctuation is proposed here; that discussion belongs to Chapter 6.

A second point on how these figures are read: the pipeline’s native plots (reproduced directly below) are indexed by raw block number rather than the normalised cycle fraction $\tau = N/N_f$ used elsewhere in this thesis. Because the eight specimens have widely different total block counts (Table 5.2), block-number comparisons across specimens should be read with each specimen’s own fatigue life in mind rather than assumed to align at a common life fraction. The shaded region at the left of each plot marks the five Phase 3 baseline blocks, common to every specimen; the boundary with Phase 4 (block 5/6) is marked with a dashed vertical line.

5.2.1. FRF Evolution

Figure 5.1 shows the broadband FRF magnitude $|H(\omega)|$, averaged across scan points, for each of the eight Group C1 specimens individually, coloured from the first Phase 3 block (blue) to the final Phase 4 block before fracture (red). This is the rawest form in which the modal evolution can be shown, before any pole extraction or feature computation, and is reported first for that reason.

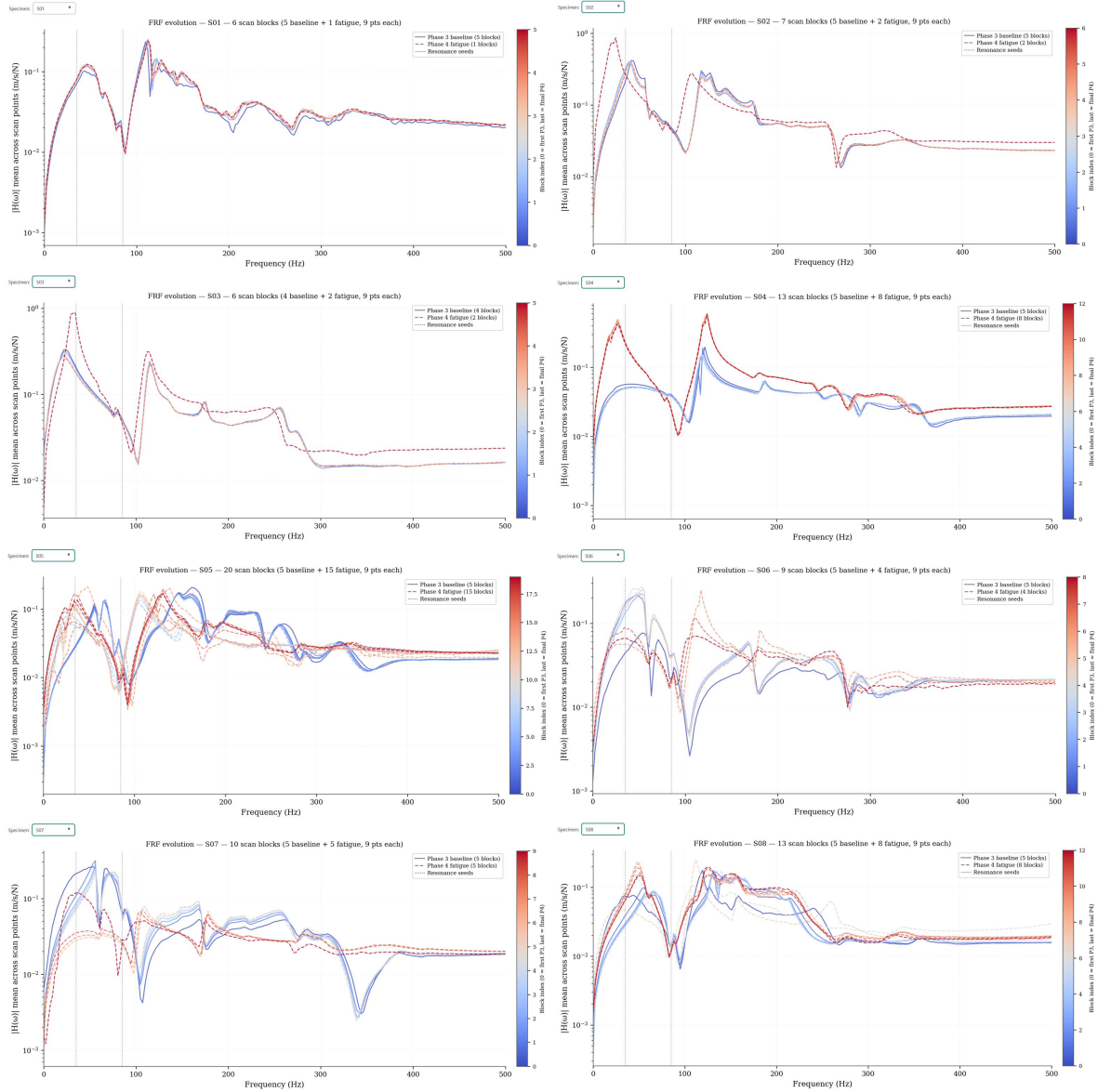


Figure 5.1: FRF magnitude evolution for each of the eight Group C1 specimens (S01–S08, corresponding to C1-01–C1-08), averaged across scan points, coloured from the first Phase 3 block (blue) to the final recorded Phase 4 block (red). Dotted vertical lines mark the resonance search-band seeds established during pole identification.

Across the eight specimens, the curves fall into three distinguishable patterns, and the distinction between them matters for how the remainder of this chapter is read.

Change confined to the final block. S01 shows the least change of any specimen in either magnitude or shape. S02 and S03 are similar for most of their length but depart abruptly in their last recorded curve: every preceding block, including the earlier Phase 4 blocks, overlays the Phase 3 baseline closely, and the departure appears only in the final block before fracture. This is the profile that damage remaining below the measurement floor until shortly before failure would be expected to produce.

A step at the phase boundary. S04, S07 and S08 show a different structure. The Phase 3 blocks form one internally tight family and the Phase 4 blocks form a second internally tight family in a different position, with the change concentrated at the boundary between the two rather than developing progressively across Phase 4. S04 is the clearest case: its Phase 4 family sits roughly half a decade higher in magnitude below 100 Hz, and its first-mode peak becomes visibly less damped, narrower and higher, without moving appreciably in frequency. In all three specimens, the blocks within each phase

overlay closely; it is between the two groups that the curves separate. Progressive fatigue damage would not be expected to hold a specimen in one stable state for five blocks, change, and then hold it in a second stable state for the remainder of its life. A change concentrated at a protocol boundary points instead to something in the measurement or loading configuration differing between the two phases. The broadband scan itself is not that difference: the shaker is reduced to a common scan level for every block in both phases (Section 3.5.1), so the scans either side of the boundary were acquired under identical excitation. What does change at the boundary is the harmonic loading between the scans, which rises from the sub-endurance Phase 3 level to the per-specimen Phase 4 level. Section 6.1.1 takes up what this implies.

Disordered evolution. S05 and S06 retain their overall FRF shape but shift to lower frequencies across the recorded range, with the shift more pronounced in the mid-frequency peaks than at the first resonance, and with successive blocks crossing one another rather than progressing monotonically in one direction.

One observation applies across the whole population. Where a shift is visible at all, it appears across most of the plotted 0–500 Hz range simultaneously rather than being isolated to one resonance while others stay fixed. A crack at the notch root is a local change in stiffness, and would be expected to affect the tracked modes by differing amounts according to how much curvature each mode carries at the notch. A shift affecting the full measured band more or less uniformly is not that signature, and is more consistent with a change in a boundary condition or in the load path, both of which act on every mode at once.

5.2.2. Frequency Shift and Anomaly Detection: Automatic Pole Selection

Figure 5.2 reports, for the automatic (MAC-based) pole-tracking path described in Section 4.1.4, four quantities together for all eight specimens: the tracked natural frequency f_n of Mode 0, the Hotelling T^2 statistic, the Harmonic Distortion Index, and the CUSUM accumulator. All four are computed from the harmonic-loading-stage signal rather than the broadband scan.

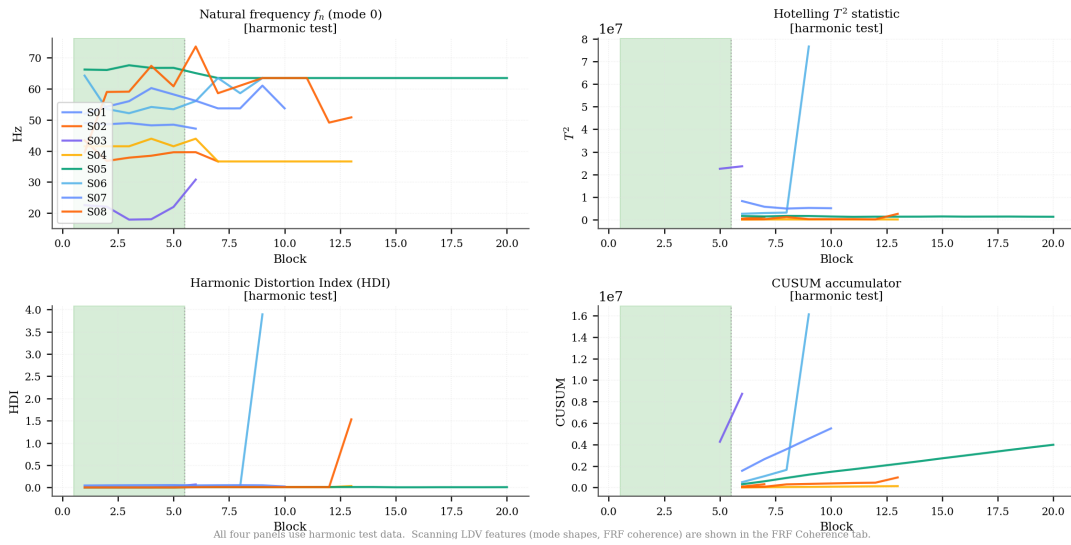


Figure 5.2: Automatic pole selection: natural frequency f_n (Mode 0), Hotelling T^2 statistic, Harmonic Distortion Index, and CUSUM accumulator, all eight Group C1 specimens, harmonic-loading-stage data. Shaded region: Phase 3 baseline blocks.

f_n stays relatively flat through Phase 3 for all eight specimens, with S08 showing more block-to-block variation than the others even within this baseline period. In Phase 4, f_n appears to plateau and hold near-constant for several blocks at a time before stepping to a new level, rather than drifting gradually. This is the same stepwise character already visible in the raw FRFs of Section 5.2.1, now expressed in the tracked pole itself: the quantity changes between stable plateaux rather than accumulating block by block, which is not the behaviour expected of progressive crack growth. T^2 and CUSUM track one another closely, as expected given CUSUM is an accumulated sum of T^2 , but CUSUM's evolution is

visibly more gradual than T^2 's sharper block-to-block excursions. HDI is flat at a low level for nearly every specimen across the full recorded range, with the exception of S06 and S08, both of which show a sudden, large-magnitude spike confined to their final block or two before fracture.

5.2.3. Frequency Shift and Anomaly Detection: Manual Pole Selection

Because pole tracking showed block-to-block fluctuation even for the two retained modes, a subset of blocks was reviewed manually using the pole-selection review workflow described in Section 4.1.5, in which the researcher inspects the stabilisation diagram directly and locks in a chosen pole by hand rather than relying on the automatic MAC-based disambiguation. Figure 5.3 shows a representative stabilisation diagram used for this review; Figure 5.4 then compares the same four quantities as Figure 5.2, but computed from the manually-reviewed poles.

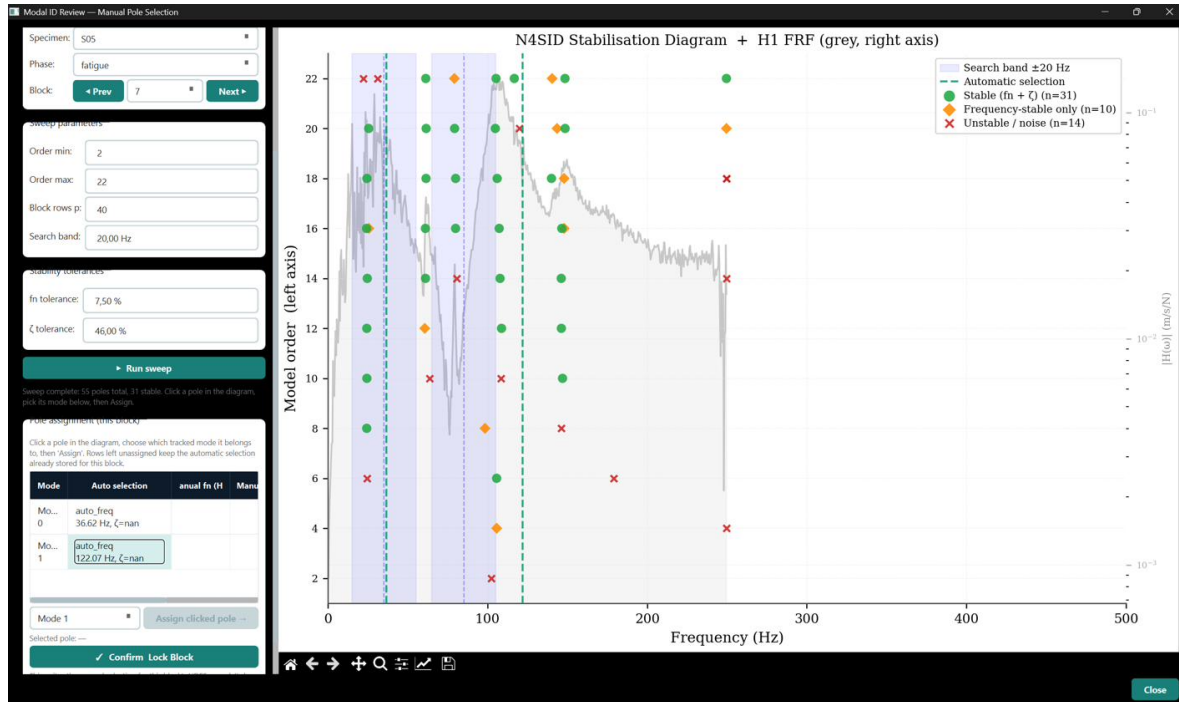


Figure 5.3: Representative N4SID stabilisation diagram used for manual pole-selection review (specimen S05, Phase 4, block 7), showing stable poles (green), frequency-stable-only poles (orange), and unstable/noise poles (red) across model orders, overlaid on the H1 FRF magnitude. The automatic selection is marked with a dashed vertical line for comparison.

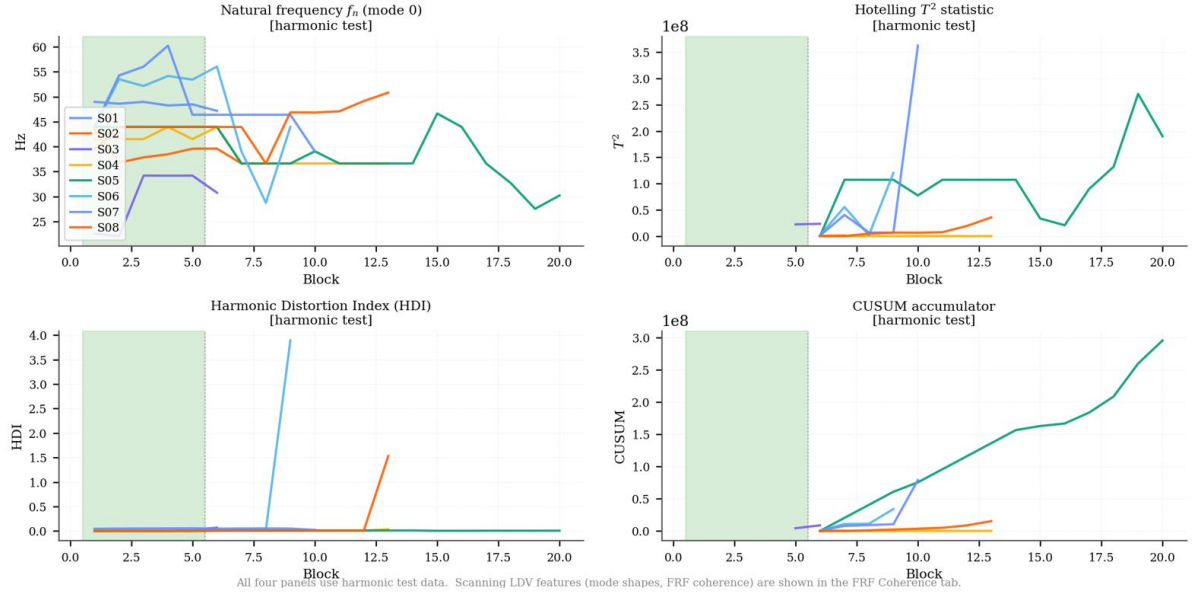


Figure 5.4: Manual pole selection: natural frequency f_n (Mode 0), Hotelling T^2 statistic, Harmonic Distortion Index, and CUSUM accumulator, all eight Group C1 specimens, harmonic-loading-stage data, poles assigned via manual review rather than automatic tracking.

Comparing Figure 5.2 and Figure 5.4 directly, the two pole-selection paths do not produce the same picture. The manually-reviewed f_n record is visibly messier than the automatic one, with more block-to-block variation for several specimens; the automatic path's tendency to plateau at a near-constant value for multiple consecutive blocks (Section 5.2.2) is less pronounced under manual review. T^2 is correspondingly less well-behaved under manual selection: it is no longer monotonically increasing for every specimen, unlike the automatic-path record. HDI and CUSUM, by contrast, show comparable behaviour between the two pole-selection paths.

5.2.4. Feature Evolution

Figure 5.5 reports the complete set of point-tier and spatial-tier features defined in Table 4.2, each computed for both tracked modes, across all eight Group C1 specimens. Each panel plots one feature (named in its title) against block number, all eight specimens overlaid in the colour coding established in Figure 5.1's legend.



Figure 5.5: Feature evolution, part 1 of 2 (Features 1–12 of 22), all eight Group C1 specimens vs. block number. See Table 4.2 for feature definitions.



Figure 5.6: Feature evolution, part 2 of 2 (Features 13–22 of 22), all eight Group C1 specimens vs. block number.

The 22 panels are grouped below by behaviour rather than reported individually. The per-mode frequency, coherence, and peak-FRF-amplitude features (both modes) show a trend distinguishable from Phase 3 scatter for most specimens from partway through Phase 4 onward, consistent with the automatic-pole-selection record of Section 5.2.2; Mode 1's versions of these features are visually more consistent across specimens than Mode 0's. The two damping-ratio features are the noisiest of the set: no specimen shows a trend clearly separable from block-to-block scatter, and damping estimation did not return a value for every block (Section 5.2.2); the exact fraction of non-convergent blocks and

whether it concentrates in particular specimens has not yet been quantified.

Phase quadrature deviation shows no trend distinguishable from noise for any specimen. Phase-diff slope and spatial phase standard deviation are both close to zero and stable for nearly the full run in every specimen except S07, which departs from this flat baseline throughout; S07's accelerometer became disconnected partway through its test, which affects these two accelerometer-derived features specifically (the FRF-based features of Section 5.2.1 are unaffected, since they do not use the accelerometer channel).

MAC and curvature-max (both modes) stay close to their expected near-baseline values for most specimens, with specimen-level exceptions: S05's Mode 0 MAC declines visibly from around block 10 onward; S06's Mode 0 MAC dips during Phase 3 itself; S02's Mode 0 and Mode 1 curvature-max both show an isolated spike within Phase 3, before Phase 4 begins. Curvature at the localised (notch) scan point shows repeated, unstructured jumps between low and high values for every specimen across the full run, with no consistent monotonic trend for any specimen. MSE-TVD (both modes) rises then falls during Phase 3, reaches a minimum at block 6 – the first Phase 4 block for every specimen – and then broadly increases through the remainder of Phase 4 for most specimens.

HDI is flat at a low, comparatively stable level for nearly the entire recorded life of every specimen, with a rise appearing only in the final block or two before fracture; the magnitude of that final-block rise varies substantially across specimens, from roughly a twofold increase to more than two orders of magnitude above the preceding plateau. Because a small number of specimens' final-block values are very large, the pooled eight-specimen panel (Figure 5.6) compresses the remaining specimens' equivalent late-life rises toward the bottom of the plotted range. Amplitude RMS and force RMS both show a step increase at the Phase 3/Phase 4 boundary, followed by a comparatively stable or mildly rising trend for most specimens, without the block-to-block reversals seen in the modal-identification-dependent features above.

5.2.5. Spatial Features: Representative Specimens

Rather than the population-level treatment used above, the spatial-tier fields – amplitude and curvature deviation at each scan point, relative to that specimen's first Phase 3 block – are reported here for two representative specimens, S04 and S05, for both tracked modes. Each panel is a heatmap with scan point index on the vertical axis and block number on the horizontal axis, coloured by the deviation in standard deviations (σ) from the reference block.

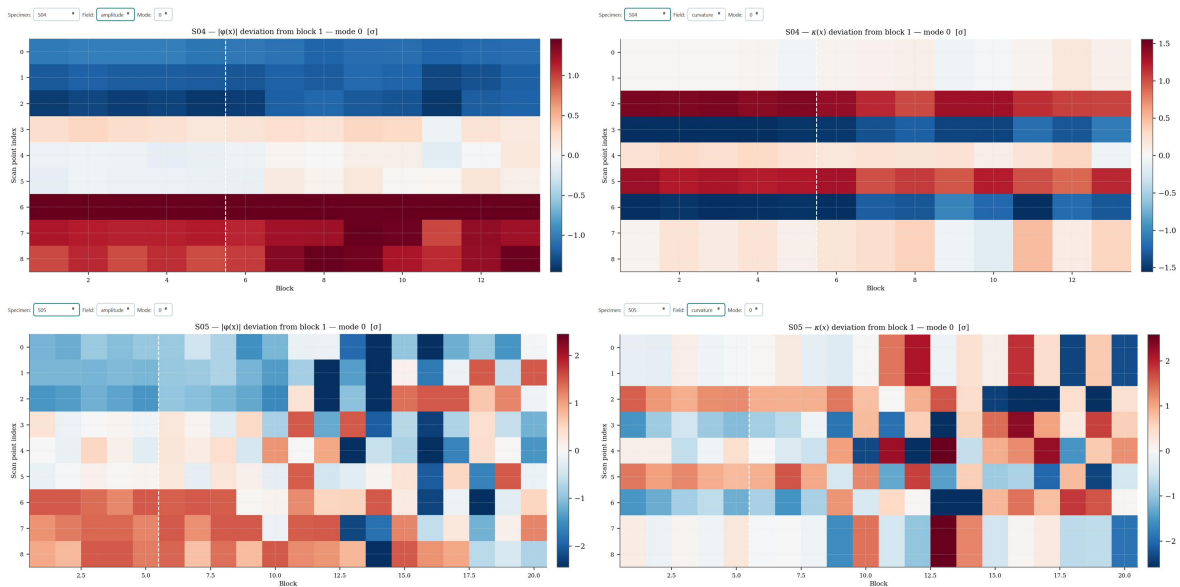


Figure 5.7: Spatial amplitude (left) and curvature (right) deviation fields, Mode 0: specimen S04 (top row) and specimen S05 (bottom row); scan point index vs. block number, coloured by deviation in σ from the first Phase 3 block. Dashed vertical line: Phase 3/Phase 4 boundary.

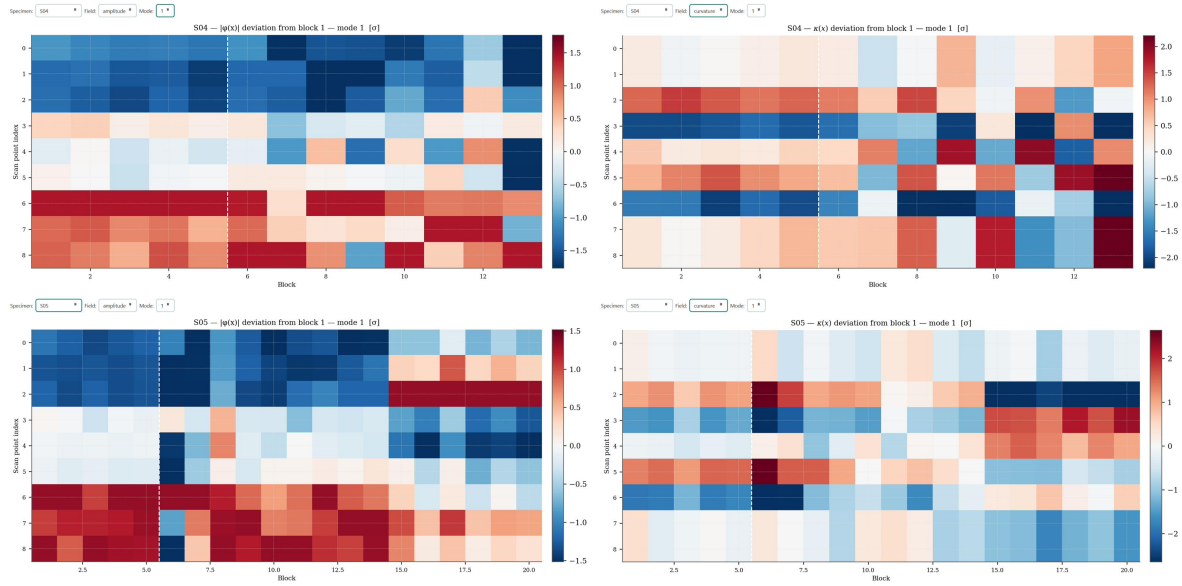


Figure 5.8: Spatial amplitude (left) and curvature (right) deviation fields, Mode 1: specimen S04 (top row) and specimen S05 (bottom row); scan point index vs. block number, coloured by deviation in σ from the first Phase 3 block.

For Mode 0 (Figure 5.7), S04's amplitude and curvature deviation fields remain fairly consistent across blocks, while S05's fields become visibly messier – less spatially structured, more block-to-block variation – toward the later blocks of its run. Comparing Figure 5.7 against Figure 5.8, the Mode 0 fields are cleaner overall than the corresponding Mode 1 fields for both specimens. Within a given specimen and mode, the amplitude and curvature deviation fields follow a broadly similar evolution.

5.2.6. Modal QA: Cross-Mode MAC Matrices

Figure 5.9 reports MAC matrices (Section 4.1.5) for a small sample of specimen and block combinations, as a check on modal identification quality independent of the damage indicators themselves. Two comparisons are shown. The block-versus-block panels compare two blocks of the same specimen directly, including two blocks drawn from the healthy Phase 3 baseline, which gives a measure of pure measurement repeatability with no damage between them. The cross-mode panels compare a fatigue block against that specimen's baseline. In both, the diagonal entries are the current-versus-baseline MAC values used as a spatial damage indicator elsewhere in this chapter, while the off-diagonal entries answer a different question: whether the two tracked modes have been kept distinct by the identification routine, or whether the mode labelled 1 in one block corresponds to the mode labelled 0 in the other.

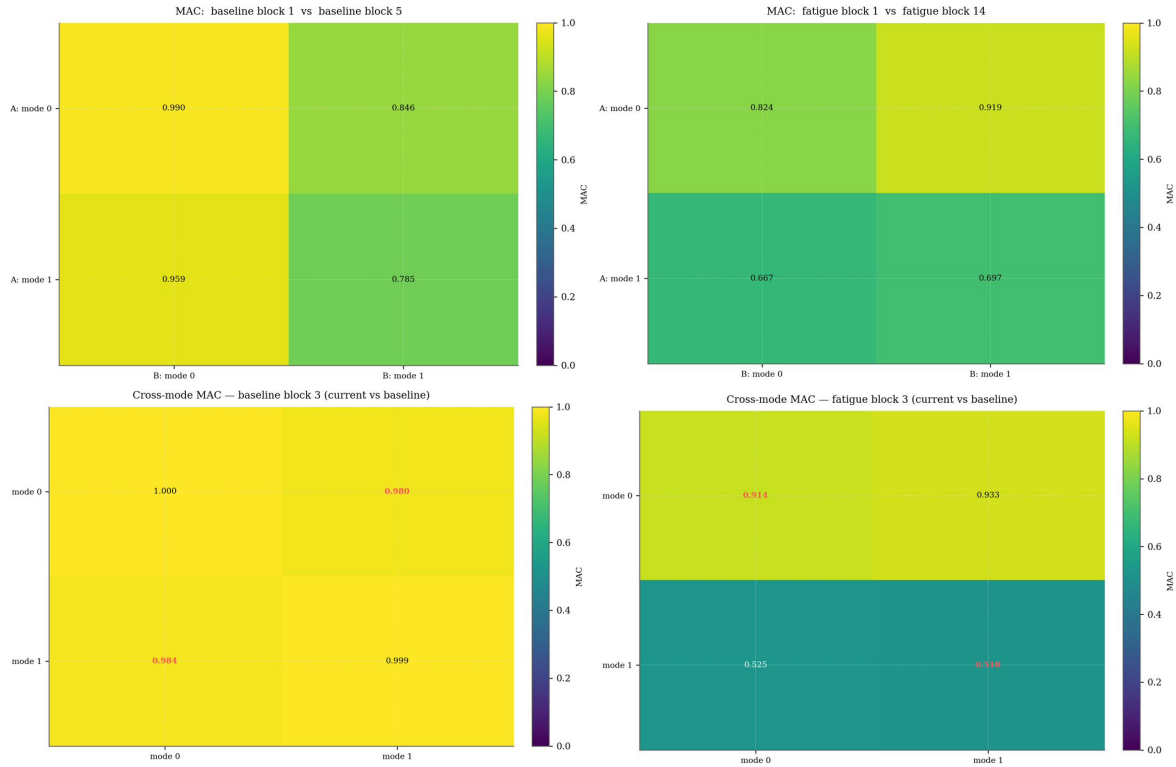


Figure 5.9: MAC matrices for a sample of specimen and block combinations, exported from the pipeline's modal QA tool. Top-left: block-versus-block comparison of two healthy Phase 3 baseline blocks of specimen S01. Bottom-right: cross-mode comparison for a Phase 4 fatigue block of specimen S06. Values beyond the 0.95 mode-mixing threshold are flagged in red by the source tool.

Two results follow from these matrices, and both bear directly on the confidence that can be placed in the mode-shape-derived indicators.

The first concerns repeatability in the absence of damage. Comparing baseline block 1 against baseline block 5 of specimen S01, two undamaged blocks of the same specimen separated only by the four intervening baseline blocks, Mode 0 returns a MAC of 0.990 as expected, but Mode 1 returns only 0.785. Since no damage accumulates during Phase 3 by design (Section 3.5.2), a Mode 1 correlation of 0.785 between two healthy states is a measurement repeatability floor rather than a damage signal. Any MAC-based indicator for that mode must resolve damage against a baseline that already varies by more than 0.2 with no damage present, leaving very little dynamic range.

The second concerns mode identity. In the same S01 baseline comparison, block 1's Mode 1 correlates more strongly with block 5's Mode 0 (0.959) than with block 5's own Mode 1 (0.785). The effect is more pronounced in the S06 fatigue-block panel, where the Mode 1 diagonal falls to 0.518 while the off-diagonal reaches 0.933: the mode labelled 1 in the fatigue block resembles the *baseline Mode 0* far more closely than it resembles the baseline Mode 1. This is not mode mixing in the sense of two distinct modes converging onto similar shapes; it indicates that the mode labels are not consistently tracking the same physical mode from block to block. That is a direct answer to the question raised in the introduction to this section, and it means Mode 1's spatial indicators cannot be assumed to describe a single physical bending mode across a specimen's life.

Every matrix reported here is 2×2 , since only two modes could be tracked with acceptable stability across the campaign (Section 5.2). The mode-identification behaviour documented here is a plausible contributor to that limit, and is taken up in Section 6.1.1.

5.3. Indicator Sensitivity Ranking (R2)

Research Question R2 asks: *which vibration-based indicators are most sensitive to early fatigue damage?* Layer 1 addresses this through the 33-feature set established in Section 5.2 (20 point-tier, 13 spatial-tier, computed for the two modes tracked with acceptable stability), using three analyses applied to the pooled feature matrix from the eight Group C1 specimens, normalised to the cycle fraction axis $\tau = N/N_f$: a Spearman rank correlation and redundancy analysis, a principal component analysis (reported here only through the PC1 loading column of the ranked table, since redundancy is already visible from the Spearman figure and a separate scree/loadings figure was not produced), and a SHAP-based ranking via a gradient-boosted regression tree, reported as two static summaries (global and early-life importance) rather than a time-resolved series (Section 5.4 note below). The primary deliverable of this section is the ranked indicator table (Table 5.3), which consolidates all three analyses into a single per-feature summary.

The two Layer 0.5 health indicators HI1 and HI2 (Section 4.2) are compared against the hand-engineered feature set in Section 5.3.4, alongside an independent leave-one-specimen-out validation of the CAE detector against the classical T^2 /Isolation Forest/CUSUM detector.

5.3.1. Damage Sensitivity and Redundancy

Figure 5.10 shows the median Spearman rank correlation coefficient $|\rho_s|$ across the eight Group C1 specimens for each of the 33 features, sorted in descending order. Redundant pairs (Pearson $|r_{ij}| > 0.90$) are shown with a hatched fill rather than a separate correlation matrix, since the pairs of interest are visually adjacent once sorted by $|\rho_s|$ (e.g. delta_mac_mode1 and mac_mode1, both at the top of the ranking).

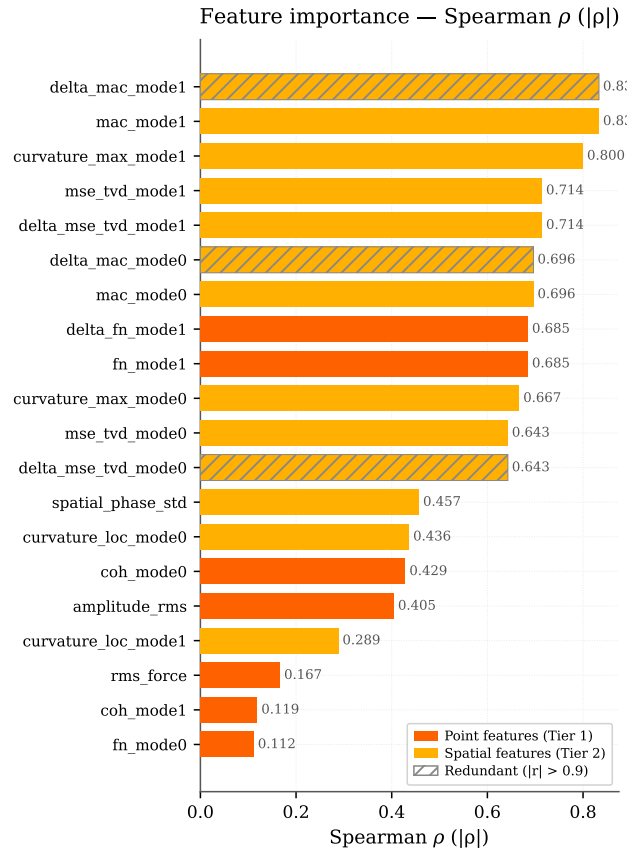


Figure 5.10: Median Spearman rank correlation $|\rho_s|$ between each of the 33 Layer 0 features and the cycle fraction τ , aggregated across the eight Group C1 specimens. Orange: point-tier (Tier 1) features; yellow: spatial-tier (Tier 2) features. Hatched bars: features flagged as redundant ($|r_{ij}| > 0.90$ with a higher-ranked partner).

The four strongest features by $|\rho_s|$ are `delta_mac_mode1` and `mac_mode1`, tied at 0.833, followed by `delta_mac_mode0` and `mac_mode0`, tied at 0.700 – all four spatial-tier, and all four MAC-derived. `curvature_max_mode1` (0.690) and `curvature_max_mode0` (0.668) follow closely, so the six most damage-correlated features in the set are all spatial-tier.

The sign structure is worth noting, since Table 5.3 reports ρ signed while the ranking uses $|\rho|$. The MAC features split into two exactly opposite pairs: `mac_mode1` at -0.833 against `delta_mac_mode1` at $+0.833$, and `mac_mode0` at -0.700 against `delta_mac_mode0` at $+0.700$. This is expected rather than informative: the delta feature is constructed as a departure from the baseline MAC, so it is the sign-inverted complement of the MAC itself, which is also why each pair is flagged redundant.

Seven features have a median $|\rho_s|$ of exactly zero and therefore do not appear in the bar chart at all: `zeta_mode0`, `zeta_mode1`, `delta_zeta_mode0`, `delta_zeta_mode1`, `phase_diff_slope`, `fn_harmonic`, and `curvature_loc_mode1`. That both damping features and both of their delta forms fall in this group is the single clearest result of this analysis: damping, which the literature review identified as one of the more promising early indicators (Section 2.5), carries no monotonic relationship with cycle fraction whatsoever in this campaign. Among the features that do vary, the weakest are `phase_quad_dev` and `delta_phase_quad_dev` (both 0.024), `hdi` and `delta_hdi` (both 0.048), and `fn_mode1` and `delta_fn_mode1` (both -0.143).

5.3.2. SHAP-Based Indicator Ranking

A gradient-boosted regression tree trained to predict τ from the pooled 33-feature vector (Section 4.3.3) yields two SHAP-derived importance summaries per feature: global importance Φ_j^{global} , averaged over the full fatigue life and shown in Figure 5.11, and early-life importance Φ_j^{early} , averaged over $\tau < 0.2$ (relaxed to $\tau < 0.4$ for specimens with fewer than 20 blocks) and shown in Figure 5.12. Features excluded from the model as redundant carry no attribution in either. As noted in Chapter 4, these are the only two SHAP quantities computed; no time-resolved $\varphi_j(\tau)$ series is available; that plot is not produced by this pipeline stage.

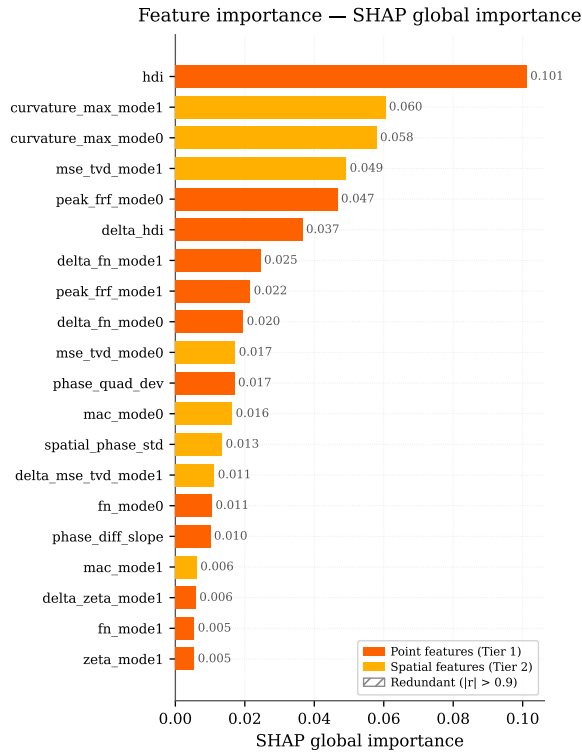


Figure 5.11: Global SHAP importance Φ_j^{global} for each of the 33 Layer 0 features, sorted in descending order. Orange: point-tier features; yellow: spatial-tier features.

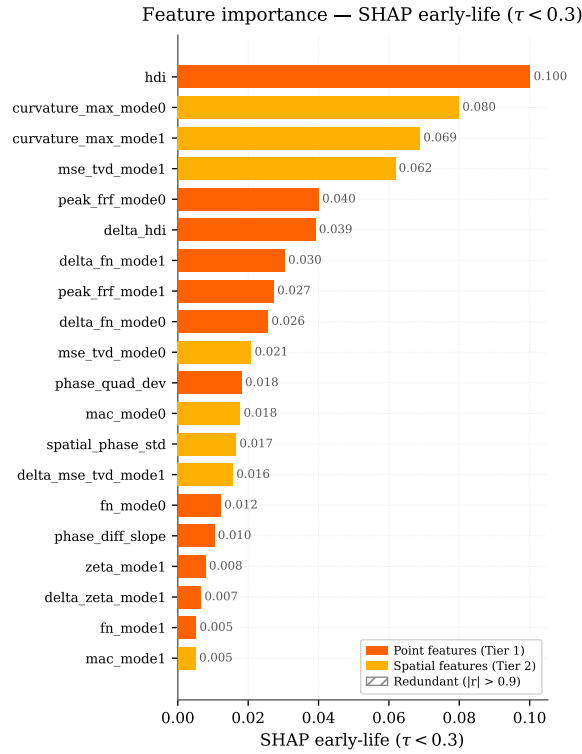


Figure 5.12: Early-life SHAP importance Φ_j^{early} ($\tau < 0.2$, relaxed to $\tau < 0.4$ for specimens with fewer than 20 blocks) for each of the 33 Layer 0 features, sorted in descending order. Orange: point-tier features; yellow: spatial-tier features.

The top five features by Φ_j^{global} are curvature_max_mode1 (0.073), hdi (0.067), delta_hdi (0.055), fn_mode0 (0.041), and delta_fn_mode0 and mse_tvd_mode1 tied at (0.039). The top five by Φ_j^{early} are hdi (0.073), curvature_max_mode1 (0.068), fn_mode0 (0.065), delta_hdi (0.051), and delta_fn_mode0 (0.044).

The composition of the top tier is therefore stable between the two views, with only the order changing: hdi and curvature_max_mode1 exchange first and second place, and fn_mode0 rises from fourth to third in the early-life view. Two observations follow. First, hdi and fn_mode0 both carry more weight early than globally (0.073 against 0.067, and 0.065 against 0.041 respectively), which is the behaviour sought from an early-damage indicator: the model relies on them most in the first fifth of life. Second, and more significantly, this ranking disagrees sharply with the Spearman ranking of Section 5.3.1. hdi has a median $|\rho_s|$ of only 0.048, placing it near the bottom of the monotonic-correlation ordering, yet it is the strongest or second-strongest feature by both SHAP measures. The MAC features invert the same disagreement: they dominate the Spearman ranking but are excluded from the SHAP model as redundant, and mac_mode1, which is retained, contributes only 0.010. A univariate monotonic correlation and a multivariate model-based attribution are measuring different things, and on this dataset they do not agree about which indicators matter.

A related caution applies to peak_frf_mode0 and peak_frf_mode1. Both carry non-zero correlation and non-zero SHAP attribution, but the underlying FRF peak magnitude does not vary monotonically across blocks: it rises and falls by varying amounts without a consistent direction, the same erratic behaviour visible in the raw curves of Section 5.2.1. Whatever these two features are contributing to the ranking is therefore unlikely to be damage.

5.3.3. Ranked Indicator Table

Table 5.3, transcribed from the pipeline's own export, consolidates the Spearman, PC1 loading, and SHAP analyses above into the primary scientific deliverable for Research Question R2: one row per feature, sorted by composite rank. The composite rank is the mean of three rank positions (Spearman

$|\rho_s|$, global SHAP, early-life SHAP), so ties in the displayed rank value are expected and legitimate wherever two features' rank-position averages coincide; sort order among tied features is resolved using $|\rho_s|$ without altering the displayed rank number (Chapter 4, Section 4.3.3).

Table 5.3: Ranked indicator table: all 33 Layer 0 features, sorted by composite rank. Spearman ρ is the median across the eight Group C1 specimens and is reported *signed*, so a feature that decreases monotonically with damage carries a negative value; the ranking itself uses $|\rho|$. SHAP columns are blank where the feature was excluded from the gradient-boosted model as redundant with a higher-ranked partner. Composite rank is the mean of the three rank positions (Spearman $|\rho|$, global SHAP, early-life SHAP), so tied and skipped rank values are expected.

Feature	Tier	Spearman ρ	SHAP glob.	SHAP early	PC1 load.	Red.	Rank
delta_mac_model	spatial	0.833	—	—	—	✓	1
curvature_max_model	spatial	0.690	0.073	0.068	—		2
delta_mac_model0	spatial	0.700	—	—	—	✓	3
fn_model0	point	0.493	0.041	0.065	0.001		5
mac_model0	spatial	-0.700	0.026	0.034	—		5
coh_model0	point	-0.671	—	—	—		6
delta_fn_model0	point	0.493	0.039	0.044	0.000		6
mse_tvd_model1	spatial	0.400	0.039	0.041	—		8
hdi	point	0.048	0.067	0.073	0.164		8
delta_hdi	point	0.048	0.055	0.051	0.986		10
curvature_loc_model0	spatial	0.447	—	—	—		11
mac_model1	spatial	-0.833	0.010	0.009	—		11
mse_tvd_model0	spatial	0.443	0.017	0.016	—		12
delta_mse_tvd_model0	spatial	0.443	—	—	—	✓	12
fn_model1	point	-0.143	0.025	0.026	0.001		12
curvature_max_model0	spatial	0.668	0.013	0.008	—		13
peak_frf_model1	point	0.238	0.017	0.024	-0.001		13
spatial_phase_std	spatial	0.457	0.007	0.011	—		14
amplitude_rms	point	0.405	—	—	—		14
coh_model1	point	-0.400	—	—	—		15
zeta_model0	point	0.000	0.020	0.020	0.001		15
zeta_model1	point	0.000	0.021	0.020	0.012		15
delta_mse_tvd_model1	spatial	0.400	0.010	0.011	—		16
delta_fn_model1	point	-0.143	0.011	0.011	0.000		17
phase_quad_dev	point	0.024	0.013	0.011	-0.002		17
peak_frf_model0	point	0.286	0.006	0.004	-0.000		19
rms_force	point	-0.167	—	—	—		20
phase_diff_slope	point	0.000	0.009	0.008	-0.000		21
delta_zeta_model0	point	0.000	0.001	0.001	0.000		23
delta_phase_quad_dev	point	0.024	—	—	—	✓	25
fn_harmonic	point	0.000	—	—	—		27
delta_zeta_model1	point	0.000	—	—	—	✓	27
curvature_loc_model1	spatial	0.000	—	—	—		27

One feature is worth flagging explicitly rather than leaving to the table: `delta_hdi` has a PC1 loading of 0.986, an order of magnitude above every other feature for which a loading is reported, the next largest being `hdi` itself at 0.164 and all others below 0.02. The first principal component of the feature matrix is therefore very nearly the change in harmonic distortion alone. This is not accompanied by a correspondingly dominant position elsewhere: `delta_hdi` sits at composite rank 10, with a median $|\rho_s|$ of 0.048 and a global SHAP importance of 0.055. Its PC1 dominance, its univariate correlation, and its model-based attribution therefore disagree with one another.

Before the ranking is read in detail, one design-phase prediction it bears on should be recorded. The Tier 3 simulations described in Section 3.1.4 of Chapter 3 predicted that a notch-localised stress concentration would produce a detectable MAC drop and curvature anomaly *before* the frequency shift alone crossed the measurement noise floor, and that prediction is what motivated the spatial-tier features being acquired at all. The Spearman ordering supports it: the four MAC features and the two maximum-curvature features occupy the six strongest positions, while the strongest frequency feature sits at composite rank 5. The SHAP ordering does not, ranking `hdi` at or near the top and placing the

retained MAC feature near the bottom. The prediction is therefore borne out on the univariate measure and contradicted on the multivariate one, and the qualification of Section 5.2.6 applies to both: the MAC repeatability floor measured between undamaged blocks limits how much of the spatial features' apparent sensitivity can be attributed to damage.

The rank-1 feature by composite rank is `delta_mac_model1`, which is itself flagged redundant with `mac_model1` (Section 5.3.1) and consequently carries no SHAP attribution at all: its composite rank rests on the Spearman component alone. The same applies to `delta_mac_model0` at rank 3. The highest-ranked feature supported by all three analyses is therefore `curvature_max_model1` at rank 2, which is the single most defensible answer this campaign offers to Research Question R2. The lowest-ranked features are `fn_harmonic`, `delta_zeta_model1`, and `curvature_loc_model1`, tied at composite rank 27; the displayed rank range runs from 1 to 27 rather than 1 to 33, since ties compress several groups of features onto a shared rank value.

5.3.4. Comparison with Layer 0.5 Health Indicators

The two Layer 0.5 autoencoder health indicators, HI1 (resonance-weighted reconstruction error) and HI2 (maximum latent z -score deviation), are reported here for a convolutional autoencoder with a 64-dimensional latent bottleneck, the primary configuration used throughout this thesis. Figure 5.13 shows both indicators' evolution across the eight Group C1 specimens, referenced against their respective Phase 3 99th-percentile thresholds.

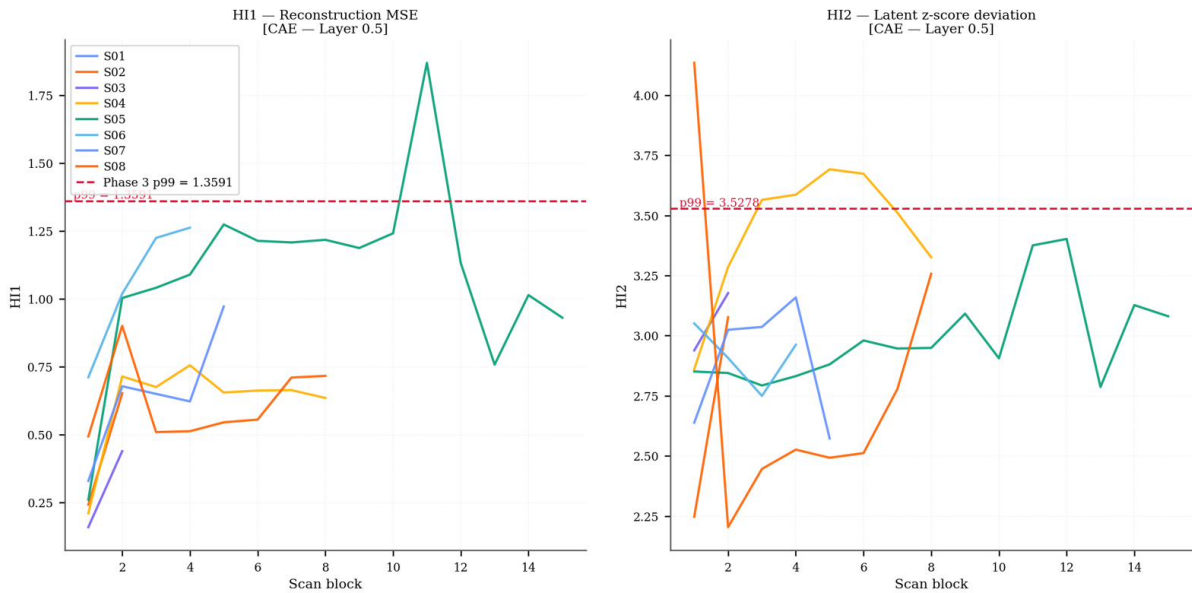


Figure 5.13: Layer 0.5 health indicators (64-dimensional bottleneck): (left) HI1, resonance-weighted reconstruction MSE; (right) HI2, maximum latent z -score deviation. Both panels show all eight Group C1 specimens vs. scan block, with the Phase 3 99th-percentile threshold shown as a dashed horizontal reference.

Against the HI1 threshold (1.359), S05 is the clearest crossing, rising well above threshold from around block 10 onward (peaking near 1.87 at block 11); no other specimen crosses the HI1 threshold convincingly, though S06 approaches it in Phase 3/early Phase 4 without clearly exceeding it. Against the HI2 threshold (3.529), S04 crosses clearly from around block 4 through block 7; S02's first recorded value is already above threshold at block 1; the remaining specimens stay below the HI2 threshold for most or all of their recorded life. The two indicators do not flag a consistent, overlapping set of specimens: S05 is the clear HI1 case but not an HI2 case, while S02 and S04 are HI2 cases without a corresponding HI1 crossing.

Beyond the health-indicator trends themselves, an independent validation was performed: a leave-one-specimen-out cross-validation (LOSO-CV) in which the CAE is trained on every specimen except one, then evaluated as a detector on the held-out specimen, compared against the independent T^2 /Isolation

Forest/CUSUM detector of Section 4.1.7 as a reference. Figure 5.14 reports the resulting ROC curves and per-specimen false-positive rates.

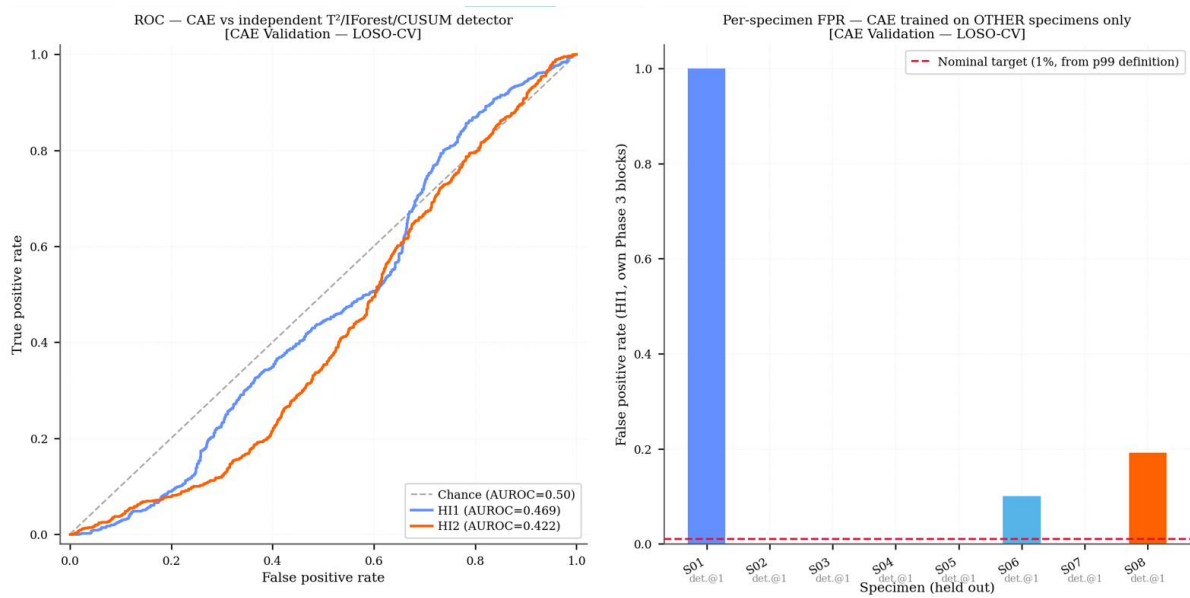


Figure 5.14: CAE validation via leave-one-specimen-out cross-validation (64-dimensional bottleneck). Left: ROC curve of HI1 and HI2 against the independent T^2 /Isolation Forest/CUSUM detector, with chance-level performance (AUROC = 0.50) shown as a dashed diagonal. Right: per-specimen false-positive rate (HI1, own Phase 3 blocks) for the CAE trained on the other seven specimens only, against a nominal 1% target derived from the 99th-percentile threshold definition.

This result is reported plainly, as the chapter’s convention requires, rather than downplayed: both HI1 and HI2 achieve an AUROC *below* 0.50 (0.469 and 0.422 respectively) against the independent detector under LOSO-CV, meaning the CAE-based detector performed worse than chance at distinguishing the reference detector’s positive and negative calls when generalising to an unseen specimen. The per-specimen false-positive rate is correspondingly uneven: specimen S01 shows a false-positive rate of 1.0 (every one of its own Phase 3 blocks flagged as anomalous by a CAE trained without it), specimens S06 and S08 show smaller but still-elevated rates, and the remaining five specimens sit at or near the 1% nominal target.

S01’s LOSO-CV false-positive-rate outlier status is not accompanied by a corresponding anomaly in its raw FRF evolution: Section 5.2.1 places S01 among the specimens with the least change in FRF magnitude or shape between Phase 3 and Phase 4 (Figure 5.1), not among the shape outliers (S06, S07) identified there.

A secondary bottleneck-dimension comparison is reported for completeness against the 64-dimensional primary configuration above: health-indicator trends for both a 16-dimensional and a 32-dimensional latent space (Figure 5.15), and LOSO-CV validation for the 32-dimensional configuration only (Figure 5.16); a 16-dimensional LOSO-CV validation was not run.

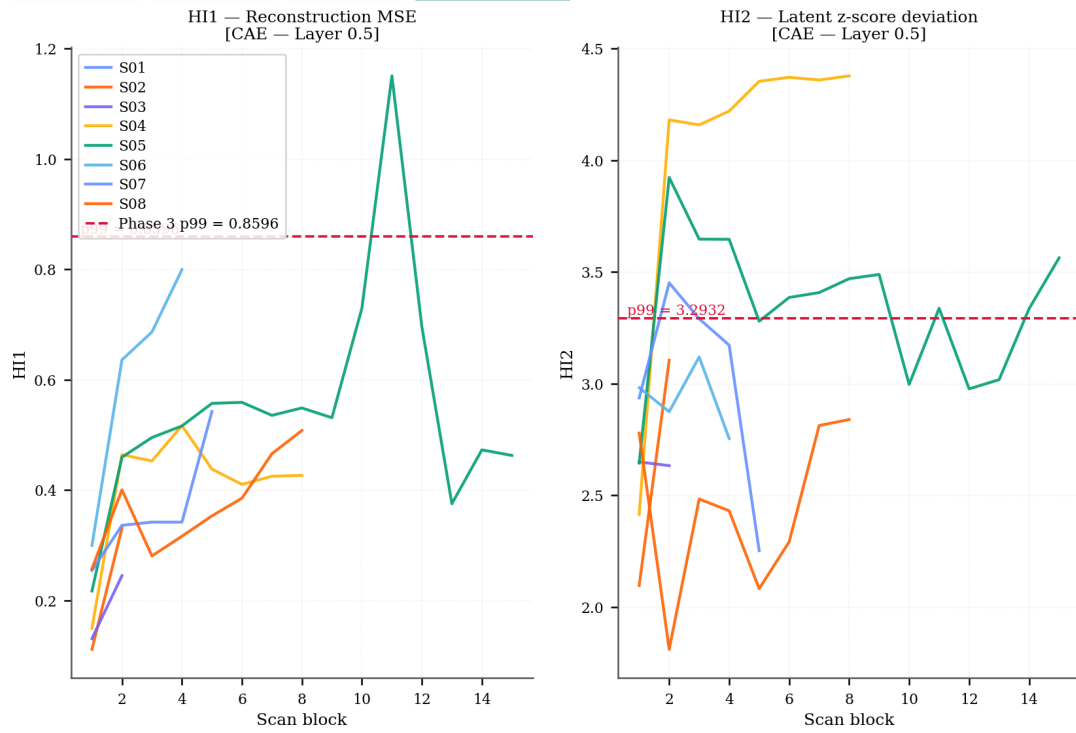
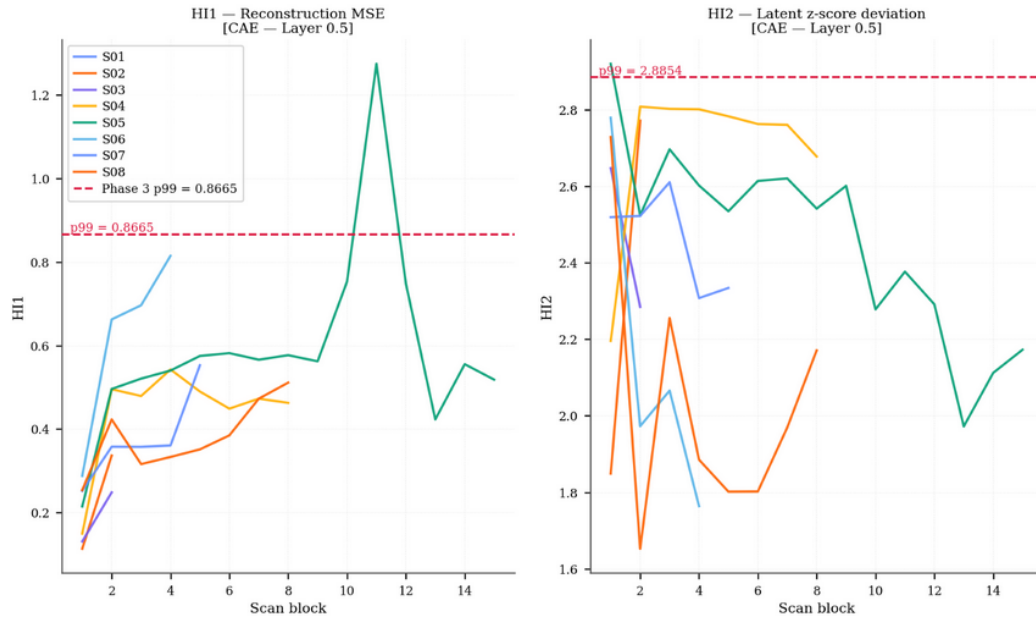
(a) 16-dimensional bottleneck, $\alpha = 5$.(b) 32-dimensional bottleneck, $\alpha = 15$.

Figure 5.15: Layer 0.5 health indicators for the 16-dimensional bottleneck ($\alpha = 5$) and 32-dimensional bottleneck ($\alpha = 15$) configurations, for comparison against Figure 5.13.

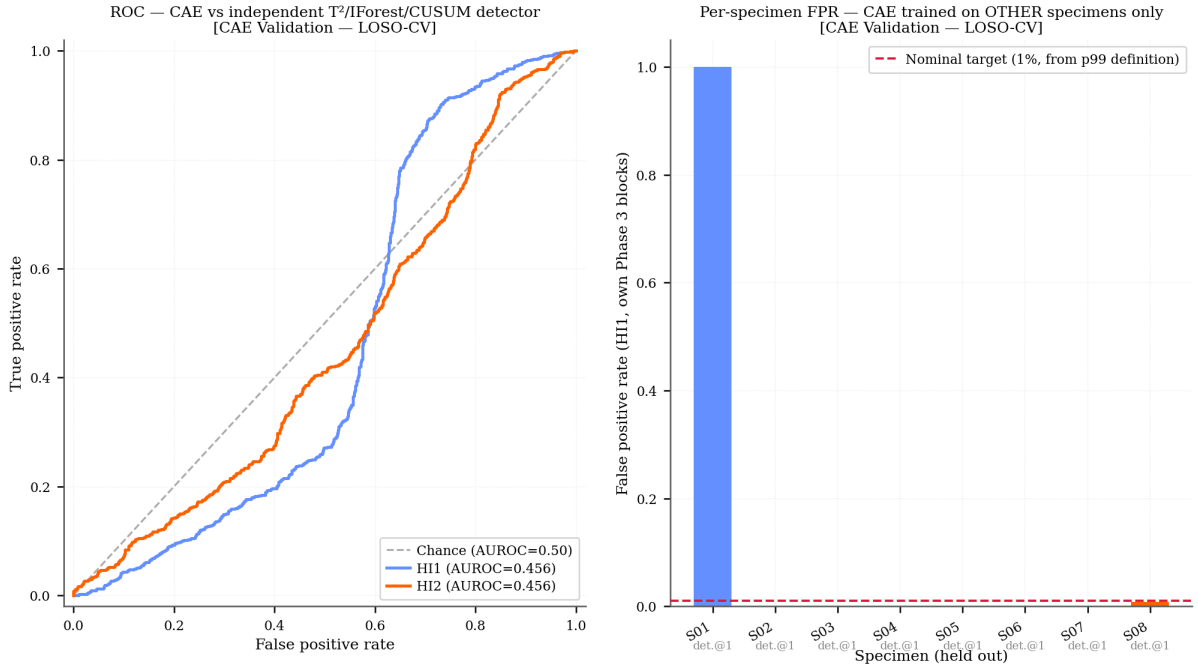


Figure 5.16: LOSO-CV validation for the 32-dimensional bottleneck configuration, for comparison against Figure 5.14.

Both the 16-dimensional and 32-dimensional health-indicator trends (Figure 5.15) show the same broad pattern as the 64-dimensional configuration: S05 is again the clearest HI1 riser, and the population otherwise stays closer to its respective threshold without a second specimen crossing as clearly. The 32-dimensional LOSO-CV validation (Figure 5.16) reproduces the 64-dimensional result closely: AUROC of 0.456 for both HI1 and HI2 against the independent detector, still below the 0.50 chance level, and S01 remains the dominant false-positive-rate outlier, with S08 again showing a smaller elevated rate and the rest of the population near the nominal 1% target.

5.4. Damage Index Trajectories (Layer 2)

Layer 2 constructs six continuous damage indices $D(\tau) \in [0, 1]$ from the Layer 0 feature record and, where available, from the Layer 0.5 autoencoder health indicators: statistical distance from the healthy baseline (D_{T^2}), stiffness loss (D_{freq}), data-driven feature consensus (D_{PC1}), SHAP-weighted sensitivity (D_{comp}), reconstruction-error deviation (D_{CAE}), and latent-space deviation ($D_{\text{CAE,latent}}$), as defined in Section 4.3.4. Computing all six in parallel rather than selecting one in advance allows this section to report, empirically, which representation of damage state is most monotonic, most consistent across specimens, and most sensitive to early change – properties relevant to the prognosis problem of Section 5.5 but not determinable from physics arguments alone.

Figure 5.17 reports all six formulations together, each specimen's raw $D(\tau)$ trajectory overlaid with its own fitted logistic S-curve (Equation 4.14 of Chapter 3), fitted independently per specimen and per formulation by nonlinear least squares against the Phase 4 data. Table 5.4 reports the resulting fit parameters, aggregated across the eight Group C1 specimens. Section 5.4.2 then compares the D_{freq} formulation specifically against the independent Synthetic Data Toolkit prediction. The two Group C2 specimens are excluded from this section; their individual $D(\tau)$ trajectories are reported in Section 5.7 as part of the independent validation track.

5.4.1. DI Trajectories and Logistic S-Curve Fits

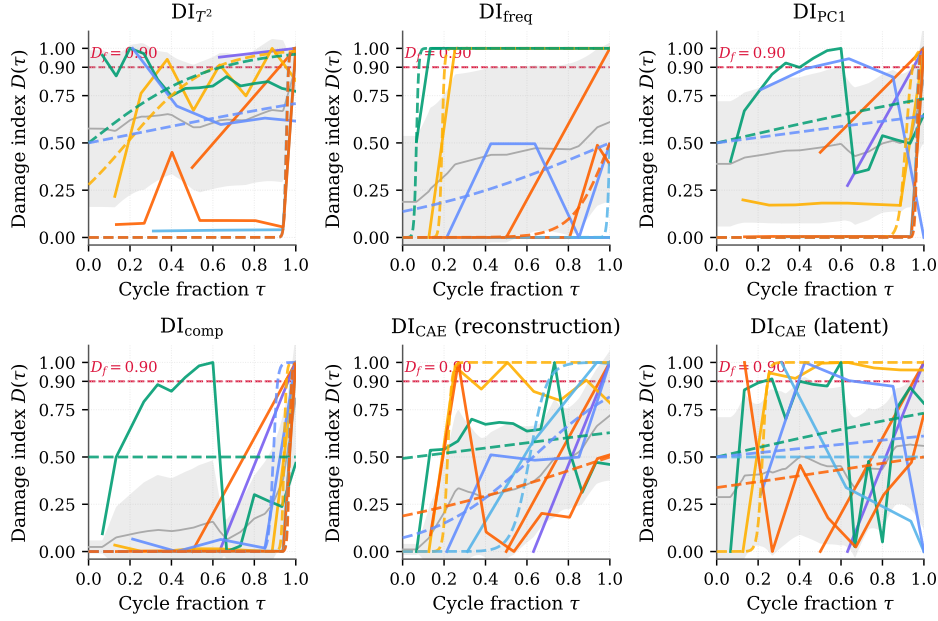


Figure 5.17: Damage index trajectories $D(\tau)$ for all six formulations, all eight Group C1 specimens (S01–S08, consistent colour coding with Figures 5.1–5.2). Solid lines: raw per-block $D(\tau)$. Dashed lines: fitted logistic S-curve per specimen. Grey shaded band: Phase 3 baseline reference range. Dotted horizontal line: failure threshold $D_f = 0.90$.

Of the six formulations, D_{comp} , D_{PC1} , and D_{CAE} (reconstruction) show the most consistently S-shaped raw trajectories across the eight specimens, with the fitted logistic curve tracking the raw per-block data closely for most specimens under these three formulations. D_{T^2} and $D_{\text{CAE,latent}}$ are visibly noisier, with several specimens' raw trajectories showing non-monotonic swings that the fitted S-curve smooths over rather than tracks point-by-point. D_{freq} 's raw trajectories are noisier still on a per-block basis than D_{comp} 's or D_{PC1} 's, consistent with the underlying frequency-shift record's own noise (Section 5.2), though a rising trend remains visible for most specimens once the S-curve is fitted. The spread in τ at which specimens cross $D = 0.5$ is narrowest for D_{comp} and widest for D_{T^2} and D_{PC1} , consistent with the $\hat{\tau}_{\text{mid}}$ IQR values reported in the fit-parameter summary below.

Table 5.4 reports the fitted $\hat{\beta}$ (steepness) and $\hat{\tau}_{\text{mid}}$ (inflection point) for all six formulations, aggregated across the eight Group C1 specimens as median (IQR), together with the number of specimens for which the fit converged.

Table 5.4: Logistic S-curve fit parameters across the eight Group C1 specimens for all six DI formulations, transcribed from the pipeline's own export. Values are median (IQR) over converged fits only. τ_{mid} : inflection point, the cycle fraction at 50% of failure-stage damage. β : steepness parameter, the damage acceleration rate.

Formulation	Converged	$\hat{\beta}$ median (IQR)	$\hat{\tau}_{\text{mid}}$ median (IQR)
Hotelling T^2 (D_{T^2})	5/8	2.001 (1.122–198.870)	0.189 (0.000–0.953)
Frequency shift (D_{freq})	5/8	61.360 (24.649–78.793)	0.119 (0.000–1.000)
PC1 score (D_{PC1})	5/8	63.097 (0.957–200.000)	0.911 (0.000–0.963)
Composite (D_{comp})	5/8	179.738 (86.218–200.000)	0.943 (0.877–0.967)
CAE reconstruction (D_{CAE})	5/8	5.104 (1.815–20.777)	0.591 (0.363–0.788)
CAE latent ($D_{\text{CAE,latent}}$)	5/8	2.350 (0.000–4.170)	0.304 (0.000–0.573)

The median $\hat{\tau}_{\text{mid}}$ is not consistent across the six formulations: it ranges from 0.119 (D_{freq}) to 0.943 (D_{comp}), spanning most of the unit interval rather than clustering around a common value. Since τ_{mid} is the cycle fraction at which damage is estimated to be accelerating fastest, the six formulations disagree about when the specimen population passes through its damage transition, some placing it in the

first eighth of life and others in the final twentieth. The median $\hat{\beta}$ varies by close to two orders of magnitude, from 2.001 (D_{T^2}) to 179.738 (D_{comp}). More informative than the medians, however, is the pattern of the interquartile bounds, which indicates that a substantial fraction of the individual fits are degenerate rather than merely imprecise. Exact round values recur across the table: $\hat{\beta}$ IQR upper bounds of precisely 200.000 for both D_{PC1} and D_{comp} , an IQR lower bound of precisely 0.000 for $D_{\text{CAE,latent}}$, $\hat{\tau}_{\text{mid}}$ IQR lower bounds of precisely 0.000 for four of the six formulations, and an upper bound of precisely 1.000 for D_{freq} . Independent nonlinear least-squares fits over different specimens do not land on identical values to three decimal places by coincidence; recurring exact values are the signature of an optimiser terminating at a parameter bound rather than at an interior optimum.

The physical readings of those bounds are unhelpful ones. A $\hat{\beta}$ at its ceiling describes a step rather than a graded transition. A $\hat{\tau}_{\text{mid}}$ at 0.000 places the damage inflection before the fatigue phase begins, and at 1.000 places it beyond the end of the record; in neither case was an inflection observed within the specimen's life. For the affected specimens the fitted parameters therefore describe the logistic form failing to find a sigmoid in the data, rather than a measured damage acceleration rate. This qualification carries into Section 5.4.2, where the D_{freq} row is compared against the Synthetic Data Toolkit's predicted inflection point. Convergence is identical across all six formulations: every one converges for exactly 5 of the 8 specimens. That the same count recurs for six independently computed formulations points to a cause in the specimens rather than in the damage representations, and the three non-converging specimens are S01–S03. These are the three shortest-lived specimens in the campaign (Table 5.2) and, having been driven at the highest excitation levels, they completed only one, two, and two Phase 4 blocks respectively before fracture. A two-parameter logistic cannot be fitted to one or two points, so the failure is one of data sufficiency rather than of the formulations: no representation of damage could have produced a fit from that record. This is a direct consequence of the block-based protocol, in which temporal resolution is inversely proportional to excitation level, and is taken up in Section 6.1.1.

D_{comp} has, by a wide margin, the tightest cross-specimen $\hat{\tau}_{\text{mid}}$ clustering of the six formulations, with an IQR width of 0.090 against 0.425–1.000 for the other five. It is also the formulation whose $\hat{\beta}$ sits closest to its ceiling, so the tight clustering describes agreement that damage rises abruptly very late in life rather than agreement about a well-resolved transition.

5.4.2. Comparison with Synthetic Data Toolkit Prediction

As described in Section 4.3.7 of Chapter 3, the Synthetic Data Toolkit produces a predicted $D(\tau)$ trajectory from a CDM–Basquin fracture mechanics model, independently of the experimental data. Rather than overlaying this prediction on the experimental $D_{\text{freq}}(\tau)$ band directly (that comparison is made narratively in Chapter 6, using the D_{freq} row of Table 5.4 against the toolkit's own predicted inflection point), Figure 5.18 instead reports the toolkit's own predicted damage evolution in isolation, for a representative excitation case: an excitation amplitude of 1.2 mm at a drive-point location 6 cm from the fixed end.

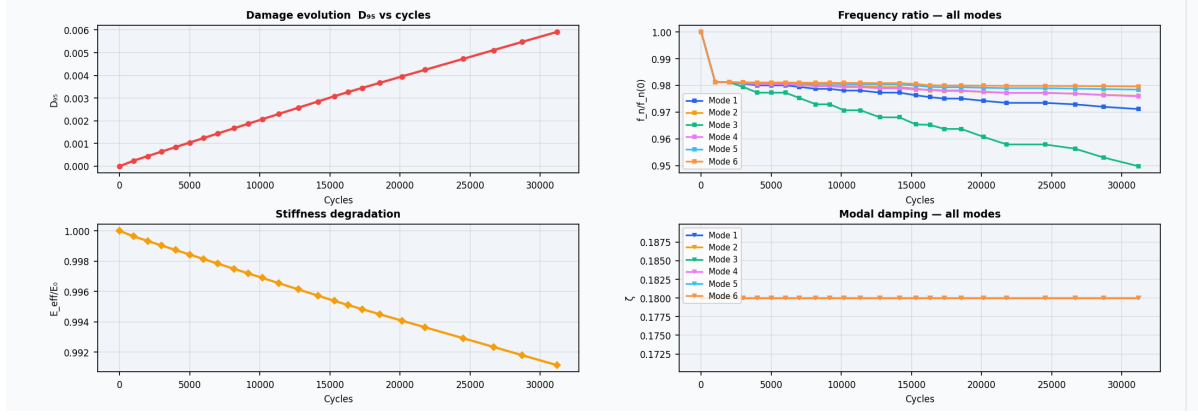


Figure 5.18: Synthetic Data Toolkit's predicted damage evolution for a representative excitation case (1.2 mm amplitude, 6 cm drive-point location). Top left: damage index D_{95} vs. cycles. Top right: natural frequency ratio $f_n/f_n(0)$ for all six simulated modes. Bottom left: effective stiffness ratio E_{eff}/E_0 . Bottom right: modal damping ratio ζ for all six modes.

This excitation amplitude is smaller than the amplitudes actually used in the Group C1 campaign (Section 3.4.3), yet the toolkit still predicts a fatigue life on the order of 35,000 cycles for this case. The predicted damage index D_{95} follows an approximately logarithmic shape as a function of cycles: rising quickly at first, then continuing to increase at a progressively slower rate. The frequency ratio decreases sharply but uniformly across all six modes over the first cycle block; beyond that initial drop, each mode continues to decline at its own, no-longer-uniform rate, with Mode 3 diverging furthest from the rest by the end of the recorded range. Stiffness degradation follows the mirror-image shape of the damage curve, as expected given the model's direct stiffness–damage coupling (Section 2.1). Modal damping is held exactly constant across all six modes for the full simulated life, a consequence of how damping is implemented in this toolkit run rather than an emergent result of the damage model.

5.5. Remaining Useful Life Estimation (R3)

Research Question R3 asks: *can vibration response trends be used to estimate remaining fatigue life?* Layer 3 addresses this through two complementary methods applied to each of the six DI formulations from Section 5.4: Method 1 (single-specimen parametric trend extrapolation, available online from the first block) and Method 2 (a population Gaussian Process, evaluated here at three checkpoints $\tau = 0.2, 0.4, 0.6$, trained on all six DI formulations). The pipeline's native output combines both methods into a single four-panel view per formulation, all eight Group C1 specimens overlaid. The three panels show: the RUL trajectory itself (true RUL, Method 1 curve and 95% CI, Method 2 checkpoint estimates with error bars), the CI width against the prognostic-horizon criterion, the raw $D(\tau)$ for reference, and a population-level summary of the prognostic horizon τ_h across all six formulations.

None of the six DI formulations – D_{T^2} , D_{freq} , D_{PC1} , D_{comp} , D_{CAE} , or $D_{\text{CAE,latent}}$ – reaches a resolved prognostic horizon at the population level within observed specimen life. Every specimen in every formulation is marked “horizon not reached” in the population summary panel. Everything else in this section should be read against that fact, and the six formulations are reported here on equal footing rather than singling one out as a working case, since none of them constitutes one.

5.5.1. RUL Trajectories: Representative Specimens

Given the population result above, no formulation is privileged as the primary case; the native Layer 3 four-panel export is shown below for all six formulations in turn, each with all eight Group C1 specimens overlaid.

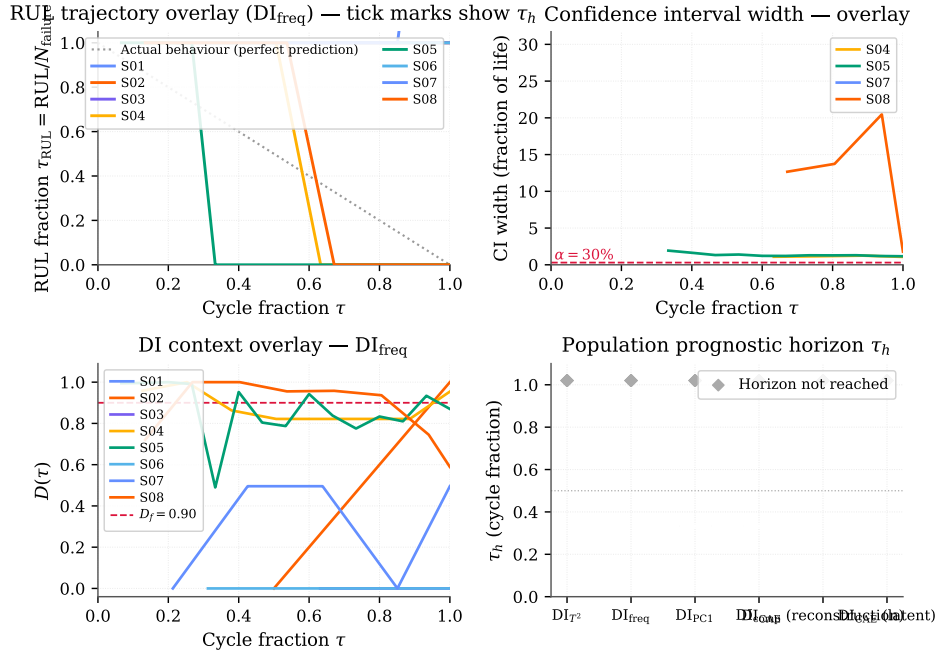


Figure 5.19: RUL trajectory (Layer 3 native export), D_{freq} formulation, all eight Group C1 specimens overlaid. Top left: RUL trajectory (true RUL, dotted; Method 1 fitted curve with 95% CI; Method 2 GP checkpoint estimates at $\tau = 0.2, 0.4, 0.6$ with error bars), tick marks show each specimen's τ_h . Top right: CI width vs. τ against the $\alpha = 30\%$ prognostic-horizon threshold. Bottom left: raw $D_{\text{freq}}(\tau)$ with failure threshold $D_f = 0.90$.

For D_{freq} (Figure 5.19), several specimens' raw $D(\tau)$ traces (bottom left) cross the $D_f = 0.90$ failure threshold, but not at a consistent τ : S05 crosses earliest, followed by S04 and S08 later in life, while S01, S03, S06, and S07 do not have a raw trace extending far enough into Phase 4 to show a crossing at all. CI width (top right) grows sharply for S08 in the last quarter of its recorded life, exceeding 20 (in fraction-of-life units) by its final block, while S05's CI width stays comparatively bounded (below 5) throughout.

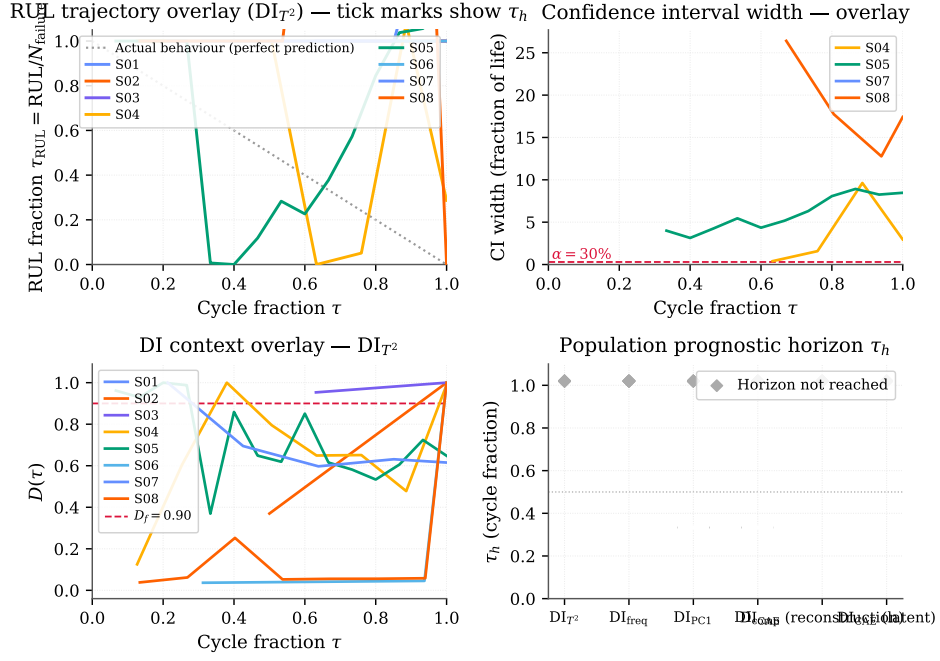


Figure 5.20: RUL trajectory (Layer 3 native export), D_{T^2} formulation, all eight Group C1 specimens overlaid. Panel layout identical to Figure 5.19.

D_{T^2} 's raw trace (Figure 5.20, bottom left) is visibly noisier than D_{freq} 's, with several specimens showing non-monotonic swings rather than a clean rise toward D_f . CI width (top right) reaches its largest recorded value for S08 (above 25), higher than any single specimen's peak CI width under D_{freq} .

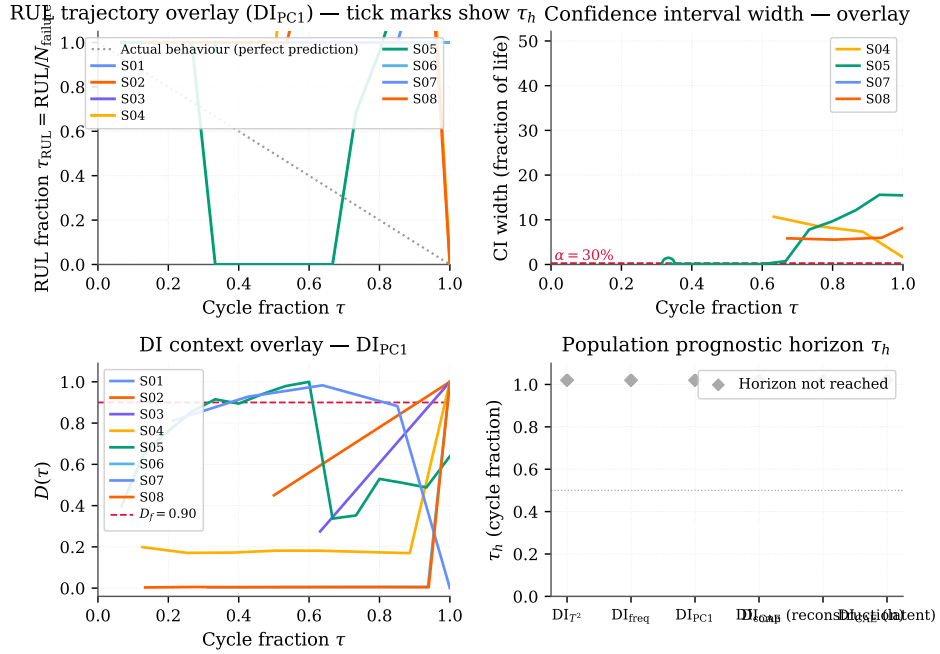


Figure 5.21: RUL trajectory (Layer 3 native export), D_{PC1} formulation, all eight Group C1 specimens overlaid. Panel layout identical to Figure 5.19.

For D_{PC1} (Figure 5.21), S05's raw trace (bottom left) is non-monotonic: it rises toward D_f , drops sharply partway through Phase 4, and partially recovers, rather than showing a single clean approach to the

failure threshold. The Method 2 GP checkpoint markers visible for other formulations are not visible on this formulation's trajectory panel (top left) for several specimens.

For D_{comp} (Figure C.1, Appendix C.1), the raw traces (bottom left) separate into two groups: S05 crosses $D_f = 0.90$ early, at $\tau \approx 0.35$, then falls back below 0.3 and rises again, while the remaining seven specimens stay near zero until approximately $\tau = 0.5$ and then climb steeply, most reaching D_f only in their final recorded block. This is the behaviour underlying the tight $\hat{\tau}_{\text{mid}}$ clustering noted in Section 5.4.1: the formulation is consistent across specimens precisely because it registers almost nothing until the end of life.

For D_{CAE} (reconstruction, Figure C.2, Appendix C.1), S04's raw trace behaves in the opposite manner, rising above D_f by $\tau \approx 0.27$ and remaining at or near saturation for the whole of the rest of its life. Its RUL estimate (top left) consequently falls to zero at $\tau \approx 0.65$ and stays there, predicting failure for the final third of a life the specimen in fact survived. S06, S07 and S08 cross D_f only in their last one or two blocks, and S05 crosses at $\tau \approx 0.73$ before falling back to 0.5.

For $D_{\text{CAE,latent}}$ (Figure C.3, Appendix C.1), no specimen exhibits a monotonic rise at all. Every trace oscillates across a substantial part of the $[0, 1]$ range for the full duration of the test, with several specimens crossing D_f and returning below it more than once, and S01 ending its life at a lower D than it began. Of the six formulations this is the one whose raw trajectories least resemble a damage accumulation process.

Two observations apply across all six formulations. The first concerns the confidence-interval panels (top right), which are the quantity the prognostic horizon criterion is evaluated against. Under D_{CAE} the maximum CI width reaches approximately 240 in units of fraction-of-life; under D_{comp} it reaches order 10^{13} , and under $D_{\text{CAE,latent}}$ order 10^{22} . A confidence interval on remaining life that is 10^{22} times the total life is not a wide estimate but a numerically meaningless one, produced by the linearised delta-method propagation of Section 4.4.4 operating on a near-singular Jacobian. The population result stated at the head of this section, that no formulation reaches a resolved prognostic horizon, is therefore only partly a statement about damage detectability: for at least three of the six formulations the $\alpha = 30\%$ criterion could not have been satisfied at any block, because the uncertainty estimate itself diverges.

The second is that only four of the eight specimens, S04, S05, S07 and S08, appear in the confidence-interval panel of any formulation. These are the four specimens with sufficient Phase 4 blocks to support the Method 1 fit; the same data-sufficiency limit that prevents S01–S03 from converging in Table 5.4 removes them from the uncertainty analysis entirely.

5.5.2. LOO-CV Prediction Accuracy

Table 5.5 reports Method 2's leave-one-out cross-validated RMSPE at checkpoints $\tau_c = 0.2, 0.4$, and 0.6 , alongside Method 1's own prediction error at the same checkpoints, for all six DI formulations. RMSPE is reported as a fraction of total specimen life ($\text{RUL}^{\text{true}}(\tau_c) = 1 - \tau_c$, Chapter 4, Section 4.4.4), consistent with how it is plotted here, regardless of the units used internally by any single computation step.

The two columns are not equivalent quantities and should not be read as a like-for-like comparison. Method 2's values are genuine leave-one-specimen-out cross-validation results: the Gaussian Process is retrained on the other seven specimens and evaluated on the held-out one, so the reported error measures generalisation to an unseen specimen. Method 1 is a stateless, single-specimen extrapolation that uses no cross-specimen information at all, so its values are in-sample prediction errors for the specimen being fitted. Method 1 will tend to look favourable at late checkpoints for that reason, and its apparent advantage at $\tau_c = 0.6$ for the two CAE formulations should be read in that light rather than as evidence of better prognostic capability.

Table 5.5: Leave-one-specimen-out cross-validated RMSPE for both remaining-life estimation methods, at three evaluation checkpoints τ_c . Method 1 produces no estimate at $\tau_c = 0.2$ for any formulation, since too few blocks are available at that point to support a parametric fit.

Formulation	Method 1 (extrapolation)			Method 2 (GP)		
	$\tau_c=0.2$	$\tau_c=0.4$	$\tau_c=0.6$	$\tau_c=0.2$	$\tau_c=0.4$	$\tau_c=0.6$
Hotelling T^2 (D_{T^2})	—	0.6000	0.1735	0.4912	0.5836	1.0896
Frequency shift (D_{freq})	—	0.6000	0.4000	0.4666	0.2899	0.5102
PC1 score (D_{PC1})	—	0.6000	0.4000	0.4092	0.7451	0.7328
Composite (D_{comp})	—	0.6000	0.4000	0.3801	0.4347	0.3572
CAE reconstruction (D_{CAE})	—	0.4012	0.0465	0.3881	0.3648	0.6213
CAE latent ($D_{\text{CAE,latent}}$)	—	0.4319	0.1146	0.3669	0.3618	0.6704

Method 1 has no reported value at $\tau_c = 0.2$ for any of the six formulations. Method 2's RMSPE does not decrease monotonically from $\tau_c = 0.2$ to $\tau_c = 0.6$ for every formulation: D_{T^2} and $D_{\text{CAE,latent}}$ both show their highest RMSPE at $\tau_c = 0.6$, the last checkpoint rather than the first. D_{comp} has the lowest Method 2 RMSPE at $\tau_c = 0.2$ (0.3801) of the six formulations. Where both methods have a reported value (at $\tau_c = 0.4$ and $\tau_c = 0.6$), Method 2 outperforms Method 1 for some formulation/checkpoint combinations and not others – for example, Method 2 is lower than Method 1 for D_{T^2} at $\tau_c = 0.4$ (0.5836 vs. 0.6000) but higher at $\tau_c = 0.6$ (1.0896 vs. 0.1735) – rather than one method consistently outperforming the other across the board.

Both DI formulations derived from Layer 0.5 (D_{CAE} , $D_{\text{CAE,latent}}$) have independently-computed Method 2 LOO-CV values in Table 5.5: an earlier implementation defect that silently restricted the GP population model to four of the six formulations has since been corrected, and the GP now trains on all six (Vol. 1 documentation, §3.5). All eight Group C1 specimens completed the Layer 0.5 autoencoder stage for this campaign, so the all-zero-DI fallback that would otherwise affect a specimen skipping Layer 0.5 (Vol. 1 documentation, §3.5) does not apply to any of the values reported here.

5.6. S-N Characterisation and Resonance Testing Context (R4)

Research Question R4 asks: *how does resonance-based vibration fatigue testing compare to traditional fatigue tests?* The S-N curve module (Section 4.5) addresses this quantitatively by fitting a Basquin stress-life model to the force records already collected for the eight Group C1 specimens and overlaying each specimen's Layer 3 prognostic horizon as a lead-time marker on the resulting S-N plot. This section reports the per-specimen force and life data first, then the Basquin fit itself, then the lead-time overlay.

The eight Group C1 specimens were tested at three distinct target *force* levels by design (Section 3.4.3), with two to three specimens per level. The stress amplitude used in the fit is $S_a = kF_a$, with k obtained from a static FEMAP finite-element analysis (Section 4.5 of Chapter 4) rather than assumed. In practice only three distinct k values were produced across the eight specimens (Table 5.6): C1-01 and C1-02 each carry their own, and the remaining six share a single value of 42.608 MPa/N. The calibration is therefore per-specimen in execution but, for three quarters of the population, a single shared factor in effect. The five Group B specimens, tested under a different protocol during the load-configuration tuning phase, are retained as contextual information for Chapter 6 rather than included in the fit.

5.6.1. FEMAP Stress Calibration

Figure 5.22 shows the static FEMAP stress field for specimen C1-03, from which that specimen's k is extracted. The remaining Group C1 specimens' FEMAP stress fields are qualitatively similar in distribution and are not reproduced individually. A separate analysis was run for every specimen, but only three distinct k values result across the eight (Table 5.6): six specimens return an identical value of 42.608 MPa/N, the geometric differences between them being too small for the linear-elastic model to resolve.

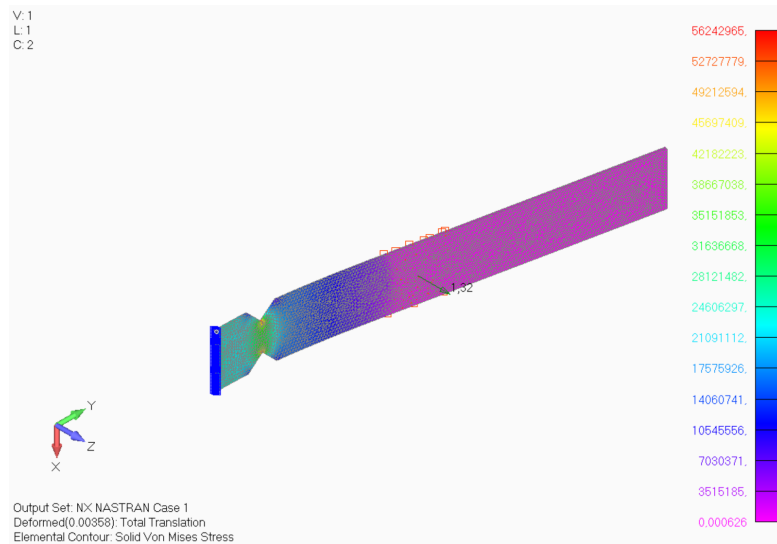


Figure 5.22: FEMAP static finite-element analysis for specimen C1-03, elemental Von Mises stress contour with exaggerated predicted displacement deformation overlaid. The notch root, where the peak stress used to derive $k_{\text{specimen}} = \sigma_{\text{FEM}}/F_{\text{FEM}}$ is extracted, is visible at the transition between the clamped section (left) and the reduced-section gauge length.

5.6.2. Per-Specimen Fatigue Life and Force Characterisation

Table 5.6 reports the per-specimen force amplitude F_a , stress ratio R , calibration factor k , resulting stress amplitude $S_a = kF_a$, and cycles to fracture N_f for all eight Group C1 specimens. F_a , R , and N_f are read directly from the Layer 3 specimen registry; N_f matches the precise pipeline-computed values in Table 5.2.

Table 5.6: Per-specimen force characterisation and fatigue life for the eight Group C1 specimens. F_a : force amplitude extracted from the first Phase 4 block. R : stress ratio. k : per-specimen FEM calibration factor (MPa/N, Section 5.6.1). $S_a = kF_a$. N_f : total Phase-4 cycles to fracture (Table 5.2).

Specimen	F_a (N)	R	k (MPa/N)	S_a (MPa)	N_f (cycles)
C1-01	4.53	−1.00	51.232	232.08	6,930
C1-02	2.87	−0.99	46.711	134.06	15,000
C1-03	2.63	−1.00	42.608	112.06	11,880
C1-04	1.97	−0.99	42.608	83.94	59,250
C1-05	4.30	−1.00	42.608	183.21	112,500
C1-06	4.55	−1.00	42.608	193.87	24,000
C1-07	4.55	−0.99	42.608	193.87	35,250
C1-08	2.57	−1.00	42.608	109.50	55,860

F_a ranges from 1.97 N (C1-04) to 4.55 N (C1-06, C1-07); the resulting S_a ranges from 83.94 MPa (C1-04) to 232.08 MPa (C1-01). N_f ranges from 6,930 cycles (C1-01) to 112,500 cycles (C1-05), a coefficient of variation of approximately 82% across the eight specimens. R is consistent across specimens at $R \approx -1$ (range −1.00 to −0.99), i.e. fully reversed loading for every specimen.

5.6.3. Basquin S-N Fit

Figure 5.23 shows the Basquin S-N fit to the eight Group C1 data points on logarithmic axes, together with the prognostic lead-time overlay reported in full in Section 5.6.4.

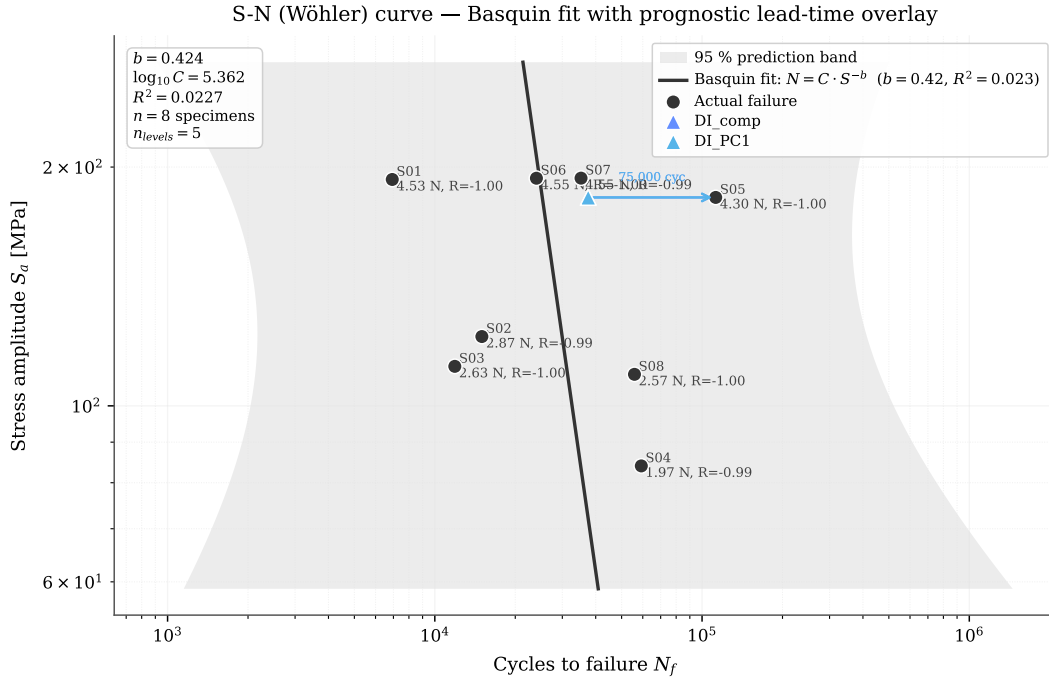


Figure 5.23: Basquin S-N characterisation of the eight Group C1 specimens (log-log axes), with the Layer 3 prognostic lead-time overlay. Filled circles: per-specimen failure data (S_a , N_f). Solid line: OLS Basquin fit $N = C \cdot S_a^{-b}$. Shaded region: 95% prediction band (individual observations). Triangles: prognostic lead-time markers for the DI formulations that reached T_h for at least one specimen; horizontal arrows point from the lead-time marker to the specimen's actual failure point, with the lead time in absolute cycles labelled. S_a values use the per-specimen k calibration (Table 5.6).

The fitted parameters are $b = 0.424$, $\log_{10}(C) = 5.362$, with $R^2 = 0.0227$ across $n = 8$ specimens. The fit's own gap rule (Section 4.5 of Chapter 4) detects $n_{\text{levels}} = 5$ distinct stress levels among the eight S_a values, rather than the three levels the campaign was designed around at the force-amplitude stage. The cause is not the calibration: because six of the eight specimens share a single k , their S_a values are directly proportional to their F_a values and preserve the force ordering exactly. Only C1-01 changes rank, moving from third to first on stress by virtue of its distinct $k = 51.232$ MPa/N. What the gap rule is detecting is instead the spread *within* each intended force level: the achieved F_a values within the nominal high-force group range from 4.30 to 4.55 N and within the mid group from 2.57 to 2.87 N, and once converted to stress these separate under the gap rule into more groups than were intended. In other words, the drive levels realised in the campaign were less repeatable than the three-level design assumed. With 5 levels detected the minimum-stress-levels warning is not triggered, the default threshold being 3, but this should not be read as a sign of fit quality: $R^2 = 0.0227$ indicates that the fitted line explains essentially none of the variance in $\log_{10}(N_f)$ across the population.

The sign of b is positive (0.424), which is the expected direction for the $N = C \cdot S_a^{-b}$ form used here, since a positive b means N_f decreases as S_a increases. Its magnitude, however, is not physically meaningful. The literature range for metallic alloys quoted in Section 2.1.1 corresponds, in the convention of Equation 4.24, to a b of roughly 7 to 20; the fitted value is between one and two orders of magnitude below that. The practical consequence is visible in the prediction itself: over the 2.76-fold stress range actually tested, from 83.94 to 232.08 MPa, an exponent of 0.424 predicts fatigue life to fall by only about a third, where the literature range would predict a fall of three to four orders of magnitude. The fitted line is therefore very nearly flat, and is better understood as the least-squares response to an uncorrelated cloud than as a calibrated material property. The 95% prediction band in Figure 5.23 spans almost the full plotted range of N_f at most stress levels, consistent with the near-zero R^2 .

The scatter driving this is visible directly in Table 5.6. C1-05, at 183.21 MPa the second-highest stress in the population, survived 112,500 cycles, the longest life recorded; C1-03, at 112.06 MPa, survived 11,880. C1-05 is also the only specimen in the campaign whose test was not completed in a single session: it was run across two days, and was therefore unclamped and remounted partway through its

fatigue life. Given the sensitivity to clamping arrangement documented in Section 5.2.1, this specimen's record should not be treated as equivalent to the other seven. Two specimens differing by a factor of 1.6 in stress differ by a factor of 9.5 in life, in the opposite direction to the one a stress-life relationship predicts. C1-06 and C1-07, tested at an identical stress of 193.87 MPa, differ from one another by 47% in life. Specimen-to-specimen variability at a fixed stress is therefore comparable to the variability across the whole stress range, which is the condition under which no S-N fit can succeed regardless of the model chosen.

5.6.4. Prognostic Lead-Time Overlay

The same Figure 5.23 carries the prognostic lead-time overlay, constructed as described in Section 4.5 of Chapter 4: each specimen's individual prognostic horizon τ_h (Section 5.5) appears as a marker at that specimen's stress level, with an arrow running to its actual failure point, so that the arrow length reads directly as lead time in cycles.

This overlay is the primary quantitative contribution to Research Question R4. Rather than leaving the comparison between resonance-based monitoring and conventional fatigue testing as a qualitative discussion point, it expresses the advance warning the pipeline provides on the same S-N axis used throughout the conventional fatigue testing literature reviewed in Section 2.3 and Section 2.10.3. Two readings are available from it in principle: comparing markers across formulations for one specimen indicates which damage mechanism dominates at that loading condition, while comparing markers across specimens at the same nominal stress level indicates whether the prognostic horizon is a property of the loading condition or of individual specimen variability. As reported below, the present campaign supports neither reading, because too few specimen and formulation combinations reached τ_h at all.

As visible in Figure 5.23, only two specimen/formulation combinations reach τ_h in this view, and both are for the same specimen: D_{comp} and D_{PC1} each reach τ_h for C1-05 only, with an identical lead time of 75,000 cycles – 66.7% of C1-05's own $N_f = 112,500$ (Table 5.6), a substantial fraction of that specimen's total life. No marker is shown for D_{freq} at any specimen in this view, and neither D_{comp} nor D_{PC1} reaches τ_h for any specimen other than C1-05. This is consistent with, not contradictory to, the population-level result of Section 5.5: none of the six formulations reached a population-level τ_h , so a single specimen reaching its own τ_h under two formulations here, however large the resulting lead time, is not itself sufficient for a population-level result – a single specimen is not a population.

D_{comp} and D_{PC1} reaching an identical lead time for the same specimen is notable on its own: both formulations are, by construction, heavily weighted toward HDI (Section 6.2.2 of Chapter 6), so an identical result for the one specimen where both reach τ_h at all is consistent with the two formulations behaving similarly whenever HDI's own signal dominates their respective trajectories, rather than an independent confirmation from two unrelated formulations. C1-05 is also the specimen whose fatigue life is least consistent with the rest of the population, being the anomalously long life at high stress that drives the near-zero R^2 of Section 5.6.3, and the only specimen tested across two separate days. The single specimen for which prognosis appears to succeed is therefore also the single specimen whose loading history is least comparable to the others. C1-05's large 75,000-cycle lead time should be read alongside Section 5.5's finding that this same specimen's raw $D_{\text{PC1}}(\tau)$ trace is non-monotonic (Figure 5.21) and its Method 2 GP checkpoint markers are not visible on the trajectory panel for this formulation – both facts are reported side by side here without resolving the apparent tension between "long lead time" and a trajectory of uncertain reliability, which belongs in Chapter 6.

5.7. Independent Validation via Optical Fractography

The two Group C2 specimens were tested under the four-stage interrupted protocol described in Section 3.5.5: pre-test optical characterisation (Stage 1), Phase 4 loading with live monitoring until the first damage alarm (Stage 2), interruption and microscope re-inspection at the alarm point (Stage 3), and resumption to final fracture followed by fracture-surface comparison against the Group C1 specimens (Stage 4). This protocol places two questions to the optical evidence:

- Q1.** At the point flagged by the live monitor as anomalous, was a visible surface change – consistent with crack initiation or early propagation – present at the notch root?

Q2. Is the fracture surface of the interrupted, re-loaded specimen qualitatively consistent with the fracture surfaces of the eight Group C1 specimens that were run to fracture without interruption?

Q1 directly tests whether the live-monitoring detection timing corresponds to a real physical event at the specimen surface, independent of the vibration-based inference. Q2 tests whether the interruption itself altered the failure mode, which is a necessary precondition for treating the Stage 3 image as representative of what a Group C1 specimen would have looked like at the same life fraction rather than as an artefact of the stop-restart loading. Both questions are answered qualitatively from the image record; no quantitative crack-length measurement or striation count is reported.

Live monitoring itself, unlike the block-wise Layer 0–3 analysis applied throughout the rest of this chapter, was exercised only in the context of these two validation runs; it was not deployed as a running system across the wider Group C1 population. It combines three of the indicators already reported elsewhere in this chapter – the Hotelling T^2 statistic, HDI, and the scanning amplitude RMS – into a live per-block dashboard, each tracked against its Phase 3-derived threshold. The two CAE-derived indicators (HI1, HI2) are computed and displayed on the same dashboard but were **excluded from the alarm-triggering logic** for both specimens, consistent with Section 5.3’s finding that the CAE-based detector achieved an AUROC below chance (0.469 and 0.422) under leave-one-specimen-out validation; the two indicators are shown in Table 5.8 for completeness only.

5.7.1. Live-Monitoring Alarm Summary

Table 5.7 reports the interruption point and total fatigue life for both specimens (consistent with Table 5.2); Table 5.8 reports the live-monitoring dashboard’s final-block indicator values against their respective thresholds for each specimen, read directly from the dashboard export (Figures 5.24–5.25, Section 5.7.2).

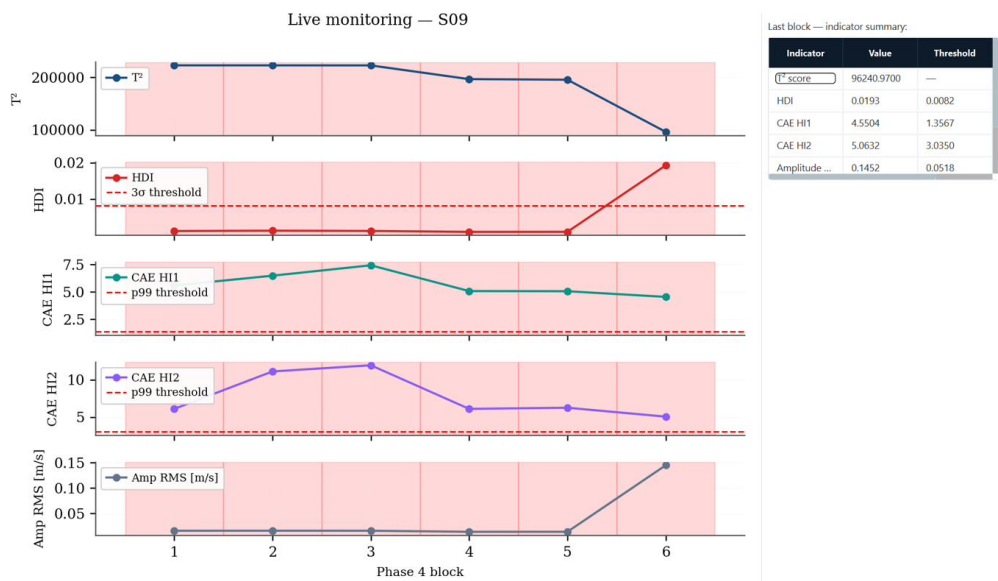


Figure 5.24: Specimen C2-01 (dashboard label S09, Table 5.7) – live-monitoring record across Phase 4 blocks up to interruption. Top to bottom: T^2 statistic; HDI (dashed line: 3σ threshold); CAE HI1 and CAE HI2 (dashed line: Phase 3 99th-percentile threshold, displayed but excluded from the alarm trigger); Amplitude RMS. Right: final-block indicator summary (reproduced in Table 5.8).

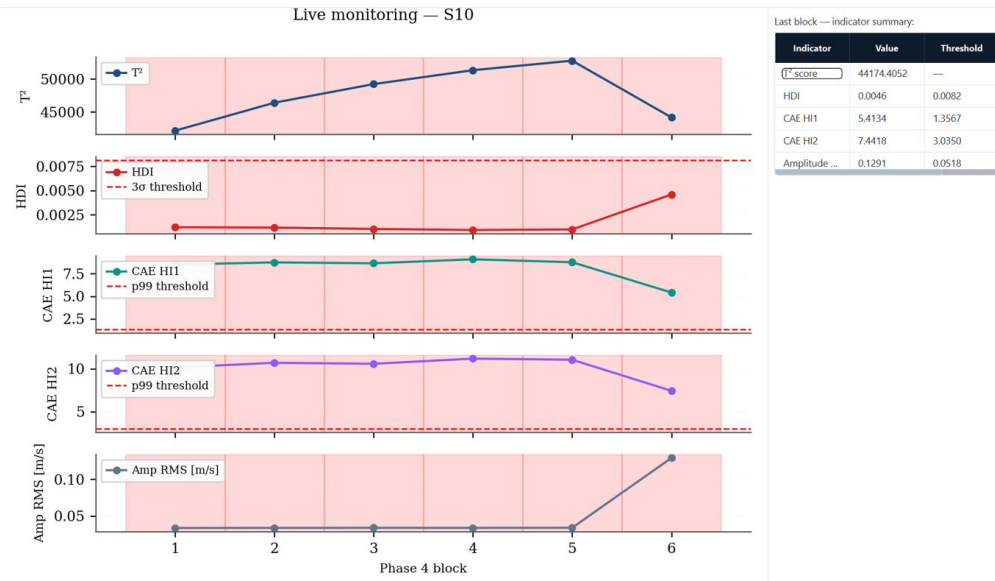


Figure 5.25: Specimen C2-02 (dashboard label S10, Table 5.7) – live-monitoring record across Phase 4 blocks up to interruption. Panel layout identical to Figure 5.24.

Table 5.7: Group C2 specimen protocol summary. $N_{\text{interrupt}}$: absolute cycle count at interruption (Phase 4 only). $N_{f,\text{total}}$: final fracture cycle count after resumption (Phase 4 only), from Table 5.2.

Specimen	Dashboard label	$N_{\text{interrupt}}$ (cycles)	$N_{f,\text{total}}$ (cycles)
C2-01	S09	7,500	$\approx 7,800$
C2-02	S10	7,500	8,200

Both specimens' interruption points and final fracture cycle counts were considerably shorter than the protocol's original target ($N_{\text{interrupt}} \approx 165,000$, $N_{f,\text{total}} \approx 245,000$), a consequence of time and equipment constraints during the campaign rather than a change to the protocol's intent.

The live-monitoring dashboard's block index (1–6) uses the same block definition as the rest of this thesis: blocks 1–5 are the standard five Phase 3 baseline blocks, and block 6 is the first (and, for both interrupted specimens, only) Phase 4 block, which is why it is the sole block entering the fatigue-life count in Table 5.7.

Table 5.8: Live-monitoring dashboard, final block before interruption: indicator values against their Phase 3-derived thresholds, for both Group C2 specimens. Bold: indicator exceeded its threshold at this block. CAE HI1/HI2 are displayed by the dashboard but excluded from the alarm-triggering logic (see text).

		C2-01 (S09)		C2-02 (S10)	
Indicator		Value	Threshold	Value	Threshold
T^2		44,174.41	—	96,240.97	—
HDI		0.0046	0.0082	0.0193	0.0082
Amplitude (m/s)	RMS	0.1291	0.0518	0.1452	0.0518
CAE HI1 (excluded)	(ex- 5.41		1.3567	4.55	1.3567
CAE HI2 (excluded)	(ex- 7.44		3.0350	5.06	3.0350

The two specimens' alarms were triggered by different combinations of indicators: C2-01's interruption was driven by Amplitude RMS alone, with HDI remaining below its threshold at the final monitored block;

C2-02's interruption was driven by both HDI and Amplitude RMS crossing their thresholds together. Neither dashboard plots an explicit T^2 threshold line, so T^2 's role in the triggering decision is not stated here as fact.

5.7.2. Stage-by-Stage Optical Evidence

Stage 1 - Pre-test characterisation

No specific $S_a/S_z/Str$ values are reported from Figure C.4 (Appendix C.2): pre-test surface roughness at the notch root remained effectively constant across the campaign and was dominated by machining imperfections rather than by anything specimen-specific, so it does not distinguish specimens or carry information relevant to the analysis. The only stage at which fracture-surface roughness differed meaningfully between specimens was Stage 4 (most visibly between the Group C1 reference specimens shown in Figure 5.30), and that difference was visible to the naked eye rather than requiring profilometer quantification.

Stage 2 - Phase 4 loading and live-monitoring alarm

Both specimens were loaded under the standard Phase 4 protocol with live monitoring active. Figures 5.24 and 5.25 show the full live-monitoring record for C2-01 and C2-02 respectively: T^2 , HDI, CAE HI1, CAE HI2, and Amplitude RMS, each plotted per monitored block.

The two specimens show the same broad pattern: across blocks 1–5 (Phase 3), all five monitored indicators are relatively flat for both specimens; at block 6 (the first, and only, Phase 4 block for both), the picture changes sharply. T^2 collapses from its Phase 3 level and HDI increases sharply for both specimens; CAE HI1 and HI2 both decrease at block 6 relative to their Phase 3 level for both specimens, the opposite direction to HDI's rise, consistent with Section 5.3's finding that the CAE-based detector is not a reliable indicator in this campaign and reported here as a fact about the dashboard record without being read as evidence for or against damage. The one point of genuine difference between the two specimens is the alarm trigger itself, reported in Section 5.7.1.

Stage 3 - Interruption and microscope re-inspection

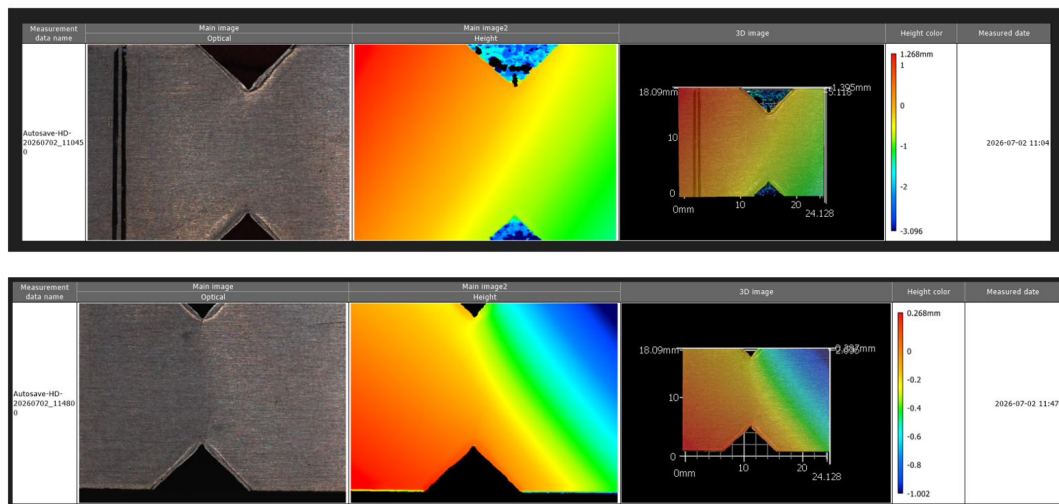


Figure 5.26: Specimen C2-01 – profilometer export of the notch-root region at Stage 1, pre-test (top row), and Stage 3, at interruption (bottom row). Both rows were acquired at the same magnification over an identical 18.09×24.128 mm field of view, and show, left to right: the optical image, the height map, the 3D reconstruction, and that row's height colour scale. The scales are auto-scaled independently per row, so colours are not comparable between them.

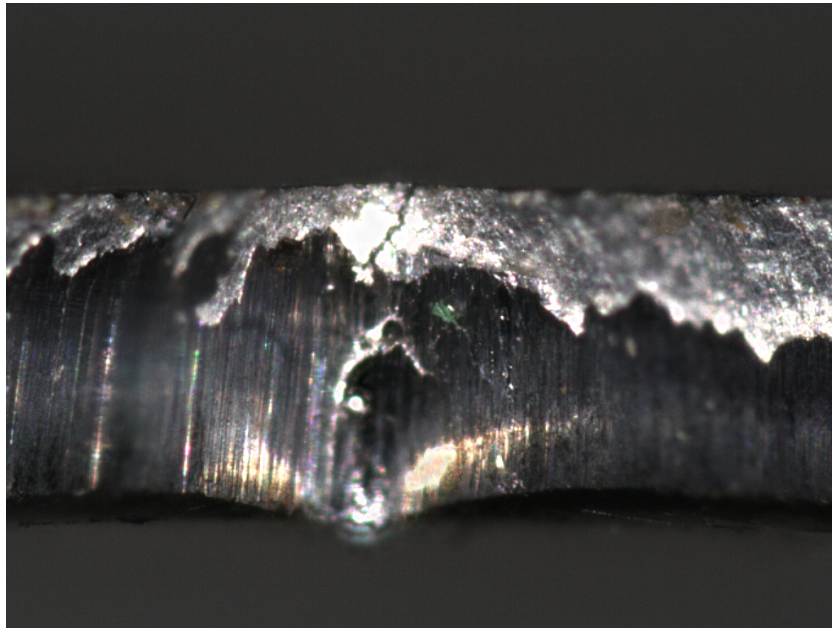


Figure 5.27: Specimen C2-01 – close-up optical image of the notch-root region at the point of interruption, showing the surface condition referenced in answering Q1.

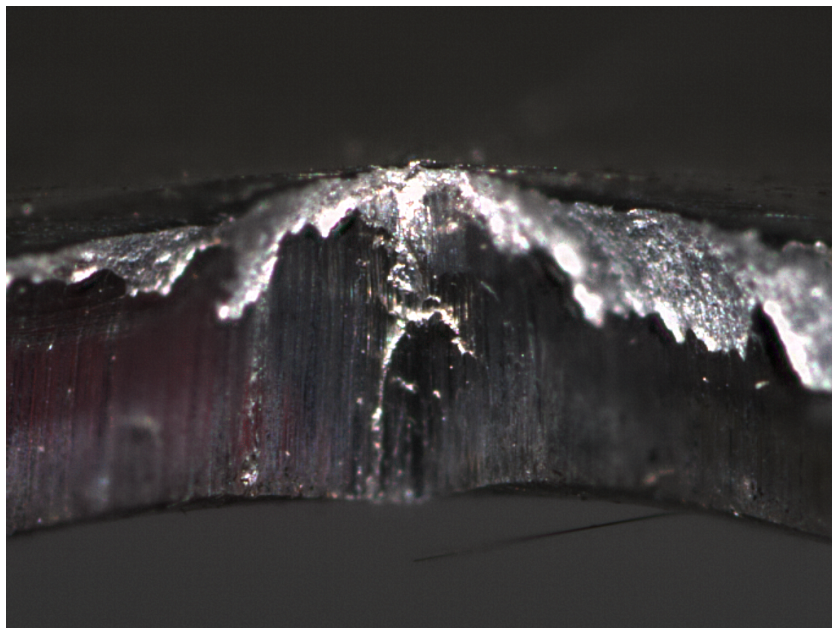


Figure 5.28: Specimen C2-02 – close-up optical image of the notch-root region at the point of interruption.

At the point of interruption, a visible surface change is present at the notch root of both specimens, already at a visually advanced, clearly identifiable stage rather than a subtle or just-initiating feature – answering Q1 in the affirmative for both specimens, and consistently between them. For C2-01 specifically (Figures 5.26–5.27): at the vertex of the notch, a crack is clearly visible, starting at the surface and extending approximately a quarter of the way down through the section; viewed from the front, two cracks are present, one propagating from each side of the notch toward the specimen's centreline.

Comparing the two rows of Figure 5.26 makes the extent of that propagation apparent. At Stage 1 the two machined notches terminate well short of one another, leaving a wide unbroken ligament across the

centre of the field; at Stage 3 the dark notch region on each side reaches substantially further toward the specimen centreline, narrowing that ligament. The additional dark area is not machined geometry but the opened crack faces. Because the two rows carry independently auto-scaled height ranges, 1.268 to -3.096 mm at Stage 1 against 0.268 to -1.002 mm at Stage 3, only the in-plane extent of the notch region should be compared between them.

For C2-02 (Figure 5.28), the notch root is imaged edge-on, through the section thickness. The upper portion of the exposed face is comparatively dark and carries fine, regular vertical striations running through the thickness, consistent with the drag marks left by the cutting stamping operation that formed the notch (Section 3.2.2) and therefore representing original machined surface. Below it, occupying roughly the lower third of the face, is a bright, coarsely textured band whose upper boundary is irregular and sharply demarcated against the darker machined region above. That boundary follows a shallow chevron profile, descending from both edges of the field toward a vertex near the centre of the section, and a brighter, approximately vertical feature rises from that vertex into the darker region above it.

The contrast in texture and reflectivity across that boundary is the crack referred to in answering Q1 for this specimen: the bright band is freshly exposed crack surface, not machined surface, and its chevron geometry converges on the section centreline in the same sense as the two cracks described for C2-01.

The two specimens' cracks are therefore alike in location, both at the notch root and both converging on the centreline, but differ in presentation: C2-01 shows two discrete cracks on an otherwise unchanged surface, whereas C2-02 shows a continuous roughened band of crack surface occupying a substantial fraction of the section. Since both close-ups were acquired in the same orientation, this difference is not an artefact of viewing geometry and reflects a genuine difference in the crack state reached at the interruption point.

What the Stage 3 evidence establishes, and what it does not.

Q1 is answered in the affirmative for both specimens: a surface change was present at the notch root at the block the live monitor flagged. Three qualifications belong alongside that answer, however, and together they bear on how much the affirmative is worth.

The first is the state of the damage at detection. The feature found at interruption was not a subtle or just-initiating one: for C2-01 the crack had already propagated approximately a quarter of the way through the section from each notch root. Detection therefore coincided with a crack of substantial extent rather than anticipating one, and the correspondence established here is between the alarm and *the presence* of damage, not between the alarm and its *onset*. Confirming the latter would require interrupting several specimens at successively earlier blocks, which the two-specimen validation track cannot do.

The second is the lead time that followed. C2-01 fractured approximately 300 cycles after resumption and C2-02 after a further 700, which on the total fatigue lives of Table 5.7 corresponds to roughly 4% and 9% of life remaining at the alarm. Whatever warning the live monitor provided, it provided it very close to failure.

The third concerns which indicator did the work. C2-01's alarm was triggered by Amplitude RMS alone, and C2-02's by Amplitude RMS together with HDI. Amplitude RMS is a monitoring channel rather than a modal damage indicator, and it places 14th of 33 in the sensitivity ranking of Table 5.3; the spatial features that dominate that ranking play no part in the live alarm, since the live monitor operates on the harmonic-loading stage and has no scan-grid data available to it. The detection reported here is therefore not a demonstration that the highest-ranked indicators of Section 5.3 detect damage, but that a single-point amplitude measure crossed a threshold at a block where damage was in fact present.

Two further features of the dashboard record run against the interpretation that block 6 was detected as a damage event on modal grounds. The T^2 statistic *falls* sharply at block 6 for both specimens rather than rising, which is the opposite of the expected direction for a statistical distance from a healthy baseline; and both CAE health indicators also decrease at block 6, in the opposite direction to HDI's rise. Neither behaviour is consistent with an anomaly detector responding to a departure from the baseline state, and both are consistent instead with block 6 differing from the Phase 3 blocks in some respect other than damage, in the manner already noted for the Group C1 population in Section 5.2.1.

Stage 4 - Final fracture and fracture-surface comparison

Following re-inspection, loading resumed for both specimens; C2-01 fractured at $N_{f,\text{total}} \approx 7,800$ cycles and C2-02 at $N_{f,\text{total}} = 8,200$ cycles (Table 5.7). Figure 5.29 shows C2-01's notch-root region in 3D height-mapped scan form across the three available stages; Figures C.5–C.6 (Appendix C.2) show both specimens' final fracture surfaces against the Group C1 reference specimens.



Figure 5.29: Specimen C2-01 – optical scans of the notch-root region at three stages: (left) Stage 1, pre-test; (centre) Stage 3, at the point of interruption; (right) Stage 4, the fracture surface after loading was resumed to failure. The right-hand panel views the fully separated fracture face edge-on, with the stable-growth region and the demarcated final-fracture zone both visible; the material surrounding it is intact beam, not an unbroken ligament. Panels are rendered with photographic surface texture rather than height colour-mapping, and are not to a common scale.

A clear bending deformation of the section is visible in the centre (Stage 3) panel of Figure 5.29, absent from the flat, intact Stage 1 geometry of the left panel, and distinct in character from the fully separated fracture texture of the right-hand (Stage 4) panel. The quantitative surface evidence for the crack itself is carried by Figure 5.26 and the close-up of Figure 5.27 rather than by this figure, which is included to show the three stages side by side. C2-02 was characterised to the same three-stage protocol, but is not reproduced as an equivalent figure here, since its progression closely mirrors C2-01's and the single detailed example is taken as representative of both specimens.

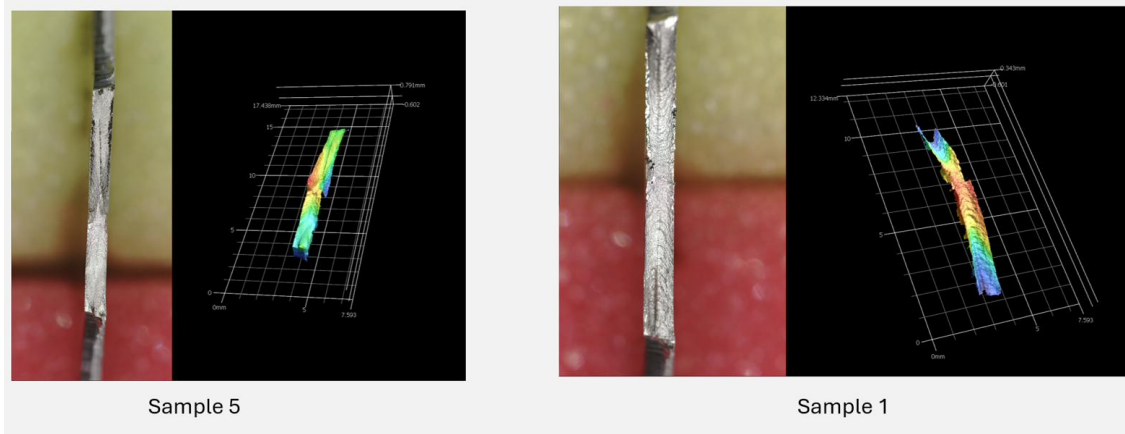


Figure 5.30: Fracture-surface comparison: Group C1 reference specimens (labelled “Sample 1” and “Sample 5” in the source export) shown alongside the corresponding 3D height-mapped scans, used as the Group C1 reference for the Stage 4 comparison against specimens C2-01 (Figure C.5 (Appendix C.2)) and C2-02 (Figure C.6 (Appendix C.2)).

Both specimens' Stage 4 fracture surfaces are qualitatively consistent with the Group C1 reference surfaces (Figure 5.30): no distinct initiation site, propagation character, or final-fracture appearance is visible that would set either C2-01 or C2-02 apart from the uninterrupted Group C1 specimens, and the two C2 specimens are, in turn, qualitatively consistent with each other. Two further, classical fractographic features are present on both surfaces, consistent with the campaign's fracture surfaces generally: a clearly demarcated final-fracture zone, distinct in character from the surrounding stable-growth region, and beach marks (macroscopic arrest-line banding). No numeric S_a roughness value was recorded for either surface: since damage was directly, visually observable at Stage 4 for both

specimens, and the fractography track is a qualitative validation by design (Section 5.7), quantifying roughness was not judged necessary.

5.7.3. Cross-Specimen Summary

Table 5.9 consolidates the key observable facts from both Group C2 specimens across the protocol.

Table 5.9: Cross-specimen summary of the Group C2 optical fractography validation. All entries are qualitative observations from the microscope image record and the live-monitoring dashboard.

Specimen	Alarm trigger	Stage 3: surface change at alarm?	Stage 4: consistent with C1?
C2-01	Amplitude RMS only	Yes – visible, advanced	Yes
C2-02	HDI + Amplitude RMS	Yes – visible, advanced	Yes

Both Group C2 specimens gave the same answer to Q1 and Q2: a visible, already-advanced surface change was present at the notch root at the point of interruption for both specimens (Section 5.7.2), and both specimens' final fracture surfaces were qualitatively consistent with the Group C1 reference specimens, with no distinct initiation site, propagation character, or final-fracture appearance setting either apart. The one point on which the two specimens differ is not Q1 or Q2, but the alarm trigger itself: C2-01's interruption was driven by Amplitude RMS alone, while C2-02's was driven by HDI and Amplitude RMS crossing their thresholds together, despite both specimens being flagged as anomalous overall. This is reported here as a factual observation carried forward for Chapter 6; whether this inconsistency undermines or is compatible with the live-monitoring detection claims is not assessed here.

5.8. Diagnostic Specimen Groups D1 and D2

Sections 5.2–5.7 report the main campaign. This section reports the two diagnostic groups introduced in Section 3.2.3, which were fabricated and tested after that campaign had concluded, for the specific purpose of examining whether the boundary condition and load transfer arrangement were responsible for the measurement behaviour documented in Section 5.2.1. Their scope is deliberately narrow. Group D1 was analysed with the LDV Analysis Suite only and Group D2 with Layer 0 only; neither contributes to the indicator ranking of Section 5.3, the damage indices of Section 5.4, the remaining-life estimation of Section 5.5, or the S-N characterisation of Section 5.6, and no D-group result is pooled with Group C1 anywhere in this thesis.

The comparison these groups permit is also narrower than it first appears, and the limits should be stated before the results. The D specimens differ from the Group C1 specimens in beam length and drive frequency as well as in clamping and load path (Section 3.2.3), and, as reported below, in manufacturing consistency. Any difference in measured behaviour between the groups therefore has several candidate causes, and the evidence here is corroborative rather than isolating.

5.8.1. Group D1: Boundary Condition and Load Transfer

Group D1's three specimens were used iteratively rather than against a fixed test matrix, in the manner of the Group B load configuration study (Section 3.4.3). Two families of arrangement were trialled in sequence: first the clamping condition, varying the clamped position and the means of seating the specimen against the fixture, and subsequently the loading arm, varying the mechanism by which shaker force is introduced into the beam. The arrangements are shown in Figure 3.11 of Chapter 3.

The clamping trials converged on a single requirement: every specimen was seated firmly against the lower face of the clamp, with the clamped length set repeatably by measuring against both a ruler and a printed reference gauge rather than by eye. The outcome of the D1 study is therefore a configuration rather than a dataset, and it is that configuration which Group D2 was then run under. No indicator trajectories, damage indices or fatigue lives are reported for Group D1, and none were computed.

5.8.2. Group D2: FRF Consistency Under the Revised Configuration

Figures 5.31 and 5.32 show the H_1 FRF, complex-averaged across the nine scan points, for the four Group D2 specimens entering the reported set, with every run for a given specimen overlaid on shared axes. Magnitude, phase and ordinary coherence are shown together, so that the quality of each estimate can be judged alongside the estimate itself. Runs are numbered for characterisation acquisitions and lettered for fatigue-loading acquisitions.

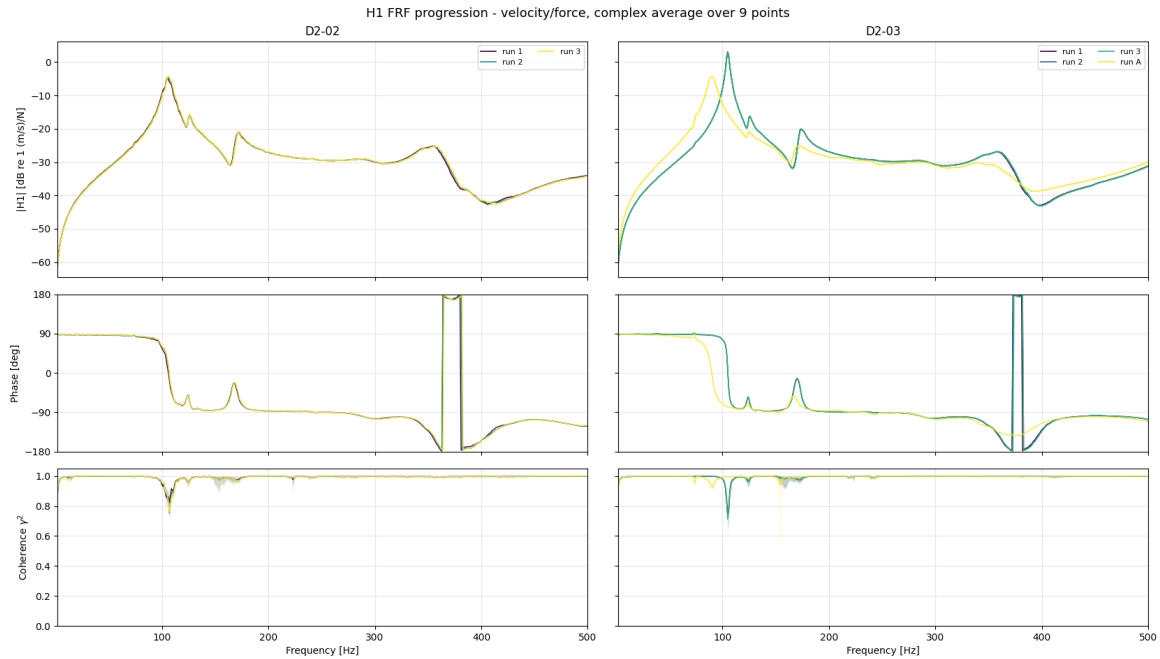


Figure 5.31: H_1 FRF progression for Group D2 specimens D2-02 and D2-03, complex-averaged over nine scan points, all runs overlaid. Top: magnitude. Centre: phase. Bottom: ordinary coherence γ^2 , with the shaded band showing the spread across scan points.

The first result is one of agreement across specimens. All four share the same FRF shape: a first bending mode near 105 Hz, a secondary peak near 125 Hz, a further resonance near 170 Hz, and an anti-resonance and mode pair between 350 and 400 Hz. The features appear at the same frequencies, in the same order, with the same relative magnitudes, in every specimen. Phase follows the same course in all four, passing through quadrature at each resonance and returning to -90° between them.

This is the clearest difference from the main campaign. Section 5.2.1 reports Group C1 specimens whose curves differ from one another not by an offset but in shape: peaks present in some records and absent in others, relative peak heights reordered between specimens, and Phase 4 families that diverge from their own Phase 3 baselines across multiple resonances at once. Under those conditions a cross-specimen comparison has no common reference to work against, which is why the population-level analyses of Sections 5.3 to 5.6 are so heavily qualified. The Group D2 records do not show that behaviour. They provide the common baseline that Group C1 does not.

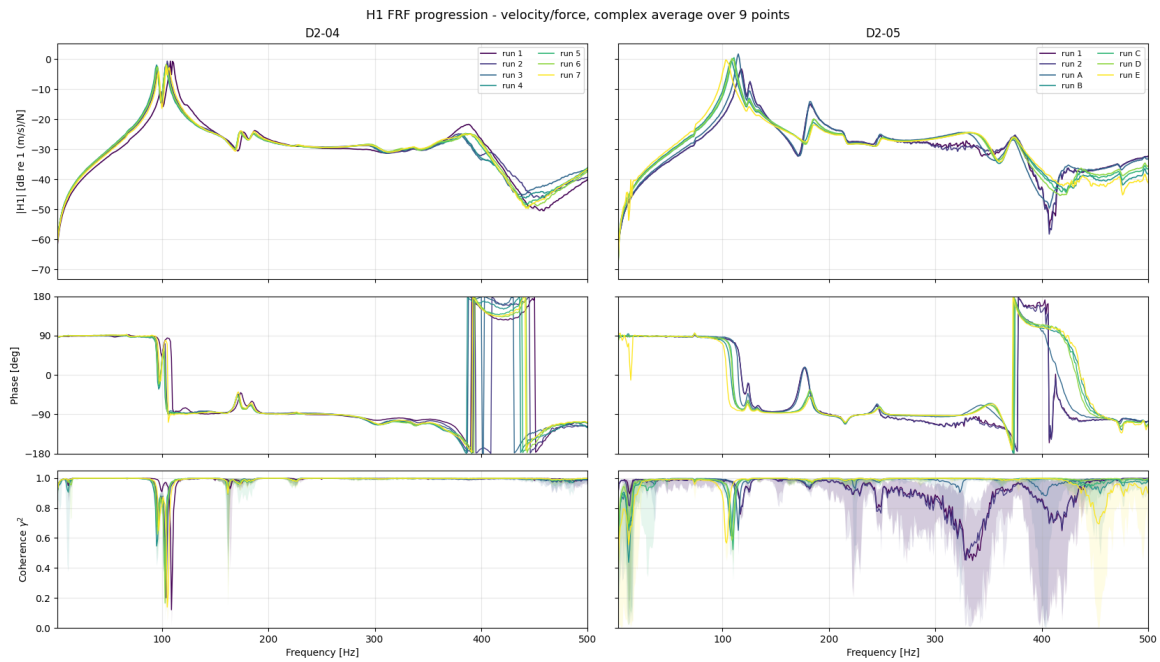


Figure 5.32: H_1 FRF progression for Group D2 specimens D2-04 and D2-05. Panel layout identical to Figure 5.31.

The second result concerns progression within a specimen. Where a specimen's runs depart from one another, they do so in an interpretable direction. D2-03 is the clearest case: its three characterisation runs overlay almost exactly, and its fatigue run sits approximately 15 Hz below them at the first mode with the mode shape and the rest of the spectrum preserved. A downward frequency shift at constant shape is the signature stiffness loss produces, and it is legible here precisely because the runs it departs from agree closely enough for the departure to stand out. D2-05, which accumulated the most fatigue loading of the four, shows the same ordering across its seven runs: the numbered characterisation runs group together and the lettered fatigue runs separate progressively from them, rather than crossing one another.

Agreement is not uniform across the four, and the exceptions are informative. D2-04's coherence falls below 0.2 in a narrow band at the first mode, indicating that the FRF estimate is locally unreliable at precisely the frequency of interest, and D2-05's coherence degrades to between 0.4 and 0.6 across the 300–400 Hz region with wide scatter across scan points. Both specimens were produced when the usual machining equipment was unavailable and are of known poorer manufacturing quality. The fixture was identical for all four, so this variation is attributable to the specimens rather than to the setup. Even in these two, however, the overall FRF shape and the ordering of characterisation against fatigue runs are preserved; what degrades is the local quality of the estimate, not the identity of the response.

The interpretation of these observations, including what they can and cannot establish about the boundary condition and load transfer diagnosis of Section 6.1.1, is taken up in Section 6.1.7 of Chapter 6.

6

Discussion

This chapter interprets the results reported in Chapter 5 in light of the literature reviewed in Chapter 2 and the methodological choices documented in Chapters 3–4. The organisation follows the four research questions – Section 6.1 (R1, modal parameter evolution), Section 6.2 (R2, indicator sensitivity ranking), Sections 6.3–6.4 (R3, damage index behaviour and remaining-useful-life estimation), and Section 6.5 (R4, resonance testing as a monitoring platform) – followed by Section 6.6, which addresses the independent live-monitoring and fractography validation evidence, and Section 6.7, which assesses the study’s broader contribution. Limitations of the study and directions for future work are consolidated in Chapter 7, where they sit alongside the per-question conclusions; result-specific caveats are raised here only where needed to qualify a specific interpretive claim.

The pipeline parameter settings under which every Chapter 5 result was produced are not repeated here as a standalone reference. Values needed to read a specific figure or table – the failure threshold $D_f = 0.90$, the prognostic horizon criterion $\Delta_{CI} < 0.30$, the 64-dimensional CAE bottleneck, the 0.90 feature-redundancy threshold, the five-block Phase 3 baseline – are already stated at their point of use in Chapter 5. Implementation-level constants that do not affect how a result is read – the T^2 confidence level, the CUSUM allowance and alarm threshold, the GP kernel choice, the CAE loss-weighting constant – are documented in Chapter 4 and are not restated here. Where a setting’s specific value is itself the subject of an interpretive argument – for instance, the per-specimen force-to-stress calibration and its effect on the S-N fit (Section 6.5), or the CAE bottleneck size and its effect on validation performance (Section 6.2) – it is discussed there, in context, rather than abstracted into a separate section here.

6.1. Modal Parameter Evolution Under Resonance Fatigue (R1)

Research Question R1 asked how the vibration response evolves as fatigue damage accumulates under resonant vibration loading. The time-series records in Section 5.2 provide the empirical answer; this section interprets those records in light of the CDM framework (Section 2.1) and the prior experimental literature on aluminium specimens under harmonic loading. Because data quality constrains what can be concluded from nearly every indicator discussed below, that constraint is addressed first, rather than left as a caveat attached retroactively to each finding.

6.1.1. Data Quality, Modal Identification Challenges, and Sources of Uncertainty

The modal parameter records reported in Section 5.2 are markedly noisier than the CDM framework’s idealised, smooth $D(\tau)$ trajectory would suggest, and this is visible before any interpretation of individual indicators is attempted. Three features of the record point the same way. The FRF evolution plots (Figure 5.1) show, for S04, S06 and S07, two internally tight families of curves either side of the Phase 3/Phase 4 boundary rather than a progression across it. The tracked natural frequency behaves the same way, holding near-constant for several blocks before stepping to a new level (Section 5.2.2). And in both Group C2 specimens the Hotelling T^2 statistic *falls* sharply at the first Phase 4 block rather than rising, which is the wrong direction for a statistical distance from a healthy baseline (Section 5.7.2).

The common element is a change concentrated at the protocol boundary. Progressive fatigue damage cannot produce that shape: it accumulates block by block, and would not hold a specimen in one stable state for five blocks, change, then hold it in a second stable state for the remainder of its life. Nor can the measurement be responsible, since the broadband scan is acquired at a common reduced excitation level in both phases (Section 3.5.1), so the scans either side of the boundary are taken under identical conditions. What changes at that boundary is the harmonic loading between the scans, which rises from the sub-endurance Phase 3 level to the per-specimen Phase 4 level. The behaviour is therefore consistent with the first application of full-amplitude loading altering something in the setup, after which the altered state persists.

Two candidate mechanisms are identified, both traceable to the experimental setup rather than to the specimens, and for both there is direct physical evidence. The first is inadequate fixity at the clamped end.



Figure 6.1: Clamping arrangement used for the Group C1 campaign, at the fixed end of the specimen. The jaw faces of the machine vice carry a stepped and grooved profile intended for gripping round or irregular stock, and scoring and edge damage from prior use are visible on the jaw surfaces.

What Figure 6.1 shows is a general-purpose clamp vice whose jaw faces carry a stepped and grooved profile, together with visible scoring and edge damage from prior use. The specimen is a flat strip one millimetre thick. Clamping a thin flat strip between profiled jaws does not produce the continuous face-to-face contact a built-in boundary condition assumes; contact is made instead along a small number of raised ridges, and the load is reacted at those discrete lines rather than distributed over an area.

Three consequences follow, and together they can account for the behaviour described above. First, the effective root of the cantilever is not at the jaw face but at whichever contact ridge happens to carry the load, so the effective cantilevered length is set by seating rather than by geometry. Since the natural frequency of a cantilever scales as $1/L^2$, a shift in the effective root of even one or two millimetres produces a frequency change of a few percent, comparable in size to the shifts being attributed to damage. Second, line contact is far more compliant than face contact, so the boundary itself contributes a non-fatigue-related compliance to every measured FRF, and that compliance depends on clamping force and on how the specimen has bedded into the grooves. Third, and most directly relevant to the step at the protocol boundary, a specimen held at discrete contact lines can settle into a new seating position when the excitation amplitude rises. It need not slip repeatedly; a single settling event at the

onset of full-amplitude loading, followed by a stable new seating, produces exactly the two-families pattern observed, and its persistence thereafter explains why the blocks within each phase agree so closely.

The wear visible on the jaw faces compounds all three, since a damaged contact surface makes the seating less repeatable between specimens as well as within a test. This is a boundary condition defined by contact mechanics that were never characterised, rather than the ideal encastre assumed by the analytical model of Section 3.1.2 and by every modal parameter extracted from the response.

The second candidate mechanism is an impure load path at the shaker attachment, for which the evidence is a fracture surface.

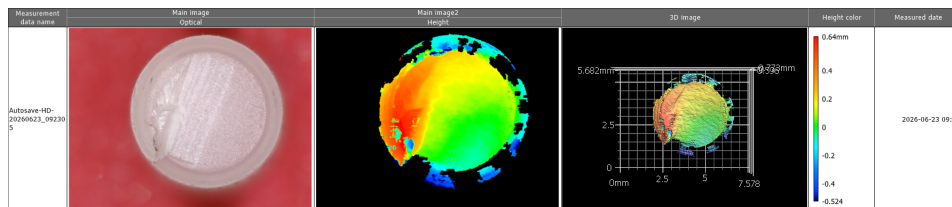


Figure 6.2: Fractured accelerometer mounting arm, Specimen 7. The fracture pattern is not consistent with a component subjected only to pure back-and-forth transverse motion, and is offered here as direct visual evidence for a mixed bending-and-axial loading history at the shaker attachment point (discussed in the text, and further in Section 6.5.1).

The accelerometer mounting arm is a slender cylindrical component whose only intended duty is to transmit motion along its own axis. A component loaded purely along its axis fractures on a plane normal to that axis, and the resulting face is flat. Figure 6.2 shows neither. The height map records a variation of approximately 1.1 mm across a face of roughly 5.7 mm diameter, so the fracture plane is inclined by something like ten degrees rather than lying normal to the axis. More tellingly, the optical view shows a single localised zone at one edge of the circular section, smoother than its surroundings and bounded by a curved front, with the remainder of the face uniformly granular. That is the classic single-origin morphology of bending fatigue: the crack initiates at the extreme fibre on the tension side, grows inward as a thumbnail, and the remaining ligament fails abruptly. Axial fatigue in a round section initiates either at an internal defect or at multiple sites around the circumference, and produces no such single edge-anchored front.

The arm was therefore carrying a bending moment, not merely an axial force, and carrying enough of one to fail in fatigue. Whatever moment the arm was reacting was also being applied to the specimen. This matters for the measurements in a specific way: the load cell reports the force along the intended axis, so any transverse or moment component is applied to the specimen but never recorded. The stress amplitude entering the S-N characterisation of Section 6.5.1 is computed from the recorded component alone, and the unrecorded component grows nonlinearly with amplitude, which is why this evidence bears on the R4 result specifically as well as on data quality generally.

Both mechanisms are amplitude-dependent, and this is what ties them to the observed step. A parasitic moment from an attachment whose geometry stiffens nonlinearly grows disproportionately as the drive level rises, and a specimen held at discrete contact lines is most likely to reseal when the load first increases. The transition from sub-endurance Phase 3 loading to full Phase 4 amplitude is precisely where both would first act, which is where the record shows the change.

The FRF evidence divides the population accordingly (Section 5.2.1 of Chapter 5). S04, S07 and S08, the three specimens whose Phase 4 curves form a distinct family from their own Phase 3 baseline, are the clearest candidates for a setup effect superimposed on any genuine damage signal. S02 and S03, which depart from baseline only in their final recorded block, are the two whose records are most readily read as damage. A further observation applies across the whole population and is difficult to reconcile with damage on its own: where a shift is visible it appears across most of the measured 0–500 Hz band at once. A notch-root crack is a local stiffness change and should affect each mode in proportion to the curvature it carries at the notch, whereas a change in boundary condition or load path acts on every mode simultaneously, which is what the record shows.

These two sources jointly explain why modal parameter identification itself, not just the subsequent damage-indicator interpretation, proved difficult. The identification challenge had two compounding causes: elevated broadband noise (consistent with the imperfect boundary condition) and closely-spaced, highly-damped modes that resisted clean separation. Of the five modal identification methods available in the pipeline, LSCF, the frequency-domain matrix-fraction method that is the conventional industrial standard, was found in practice to produce ambiguous stabilisation diagrams for this campaign's baseline FRF measurements, with genuine structural poles and spurious noise-driven poles not cleanly separable across the tested model-order range. This is the expected failure mode for a frequency-domain global fit under precisely the two conditions present here, elevated broadband noise and closely-spaced heavily-damped modes, and it is why the time-domain subspace method N4SID was adopted as the working method instead (Section 4.1.3 of Chapter 4). Appendix B sets out the mechanism of that failure and the reasons N4SID does not share it.

The exact fraction of blocks or specimens for which damping estimation returned NaN or was flagged non-convergent, and whether this was concentrated in particular specimens or spread evenly across the population, was not computed for this campaign; this is stated here, and in Section 5.2 of Chapter 5, as a known, permanent gap in the record rather than a figure still to be supplied.

Even with N4SID as the working method, the two tracked modes showed block-to-block pole fluctuation beyond what a well-conditioned identification would produce, and the MAC diagnostic of Figure 5.9 establishes two things about the consequences.

The first concerns the measurement floor. Two *healthy* baseline blocks of the same specimen, between which no damage accumulates by design, return a Mode 1 shape correlation of only 0.785 (Section 5.2.6 of Chapter 5). That is a repeatability floor, not a damage signal, and it means every MAC-based indicator for that mode is attempting to resolve damage against a baseline that already varies by more than 0.2 with nothing happening. This is the quantitative reason the spatial indicators are harder to trust than their high ranking in Section 5.3 suggests, and it applies to the undamaged state, so no amount of damage accumulation improves it.

The second concerns mode identity, and is more serious. In both the healthy comparison and the fatigue-phase panel, the mode labelled 1 correlates more strongly with the *baseline* Mode 0 than with baseline Mode 1. This is not mode mixing in the sense of two distinct modes converging onto similar shapes; it means the labels are not consistently tracking the same physical mode from block to block. Any per-mode trajectory in this thesis is therefore a trajectory of whatever the identification routine assigned that label in each block, which is a weaker object than it appears. Two consequences follow for the rest of this chapter. A discontinuity in a per-mode indicator cannot be distinguished from a relabelling event without independent confirmation, and the CDM ordering of Section 6.1.2, which is defined per mode, cannot be tested against these trajectories as cleanly as the framework assumes.

Both findings rest on a four-panel sample, so the frequency with which label swapping occurred across the full campaign is not established. That it occurs at all, and occurs between two undamaged blocks of the same specimen, is sufficient for the conclusions drawn here. A denser sampling of comparison blocks is a concrete, low-cost improvement for any repeat of this campaign.

6.1.2. The Expected Sequence: CDM Framework Predictions

The CDM framework predicts a specific ordering in which the tracked indicators should respond to accumulating fatigue damage, set out in Section 2.4.3 of Chapter 2 and illustrated in Figure 2.2: non-linear signatures first, during crack initiation, as crack-face contact generates super-harmonics; then phase deviation and damping increase; then natural frequency decline during propagation, as stiffness loss accelerates; and spatial indicators (MAC, CDI, MSE-TVD) last, as mode shape distortion localises near the crack immediately before failure. That prediction is qualitative, and the stage boundaries of Figure 2.2 are illustrative rather than calibrated, so what follows tests the *ordering* rather than the timing. Section 6.1.1 above establishes why this ordering cannot be assessed cleanly from the modal-identification-dependent indicators (frequency, damping, coherence, peak amplitude) in the way the framework anticipates: noise and identification difficulty corrupt exactly the indicators the CDM ordering leads with. The indicators that remain interpretable are, not coincidentally, the ones that do not depend on modal parameter extraction at all – HDI, amplitude RMS, and force RMS – and it is against

these, rather than against frequency or damping directly, that the CDM-predicted early-sensitivity claim can still be tested with this data.

6.1.3. Frequency, Damping, Coherence, and Peak Amplitude: The Cost of Modal Identification Dependence

Frequency shift, the CDM framework's primary damage-tracking variable (Section 2.1), proved unreliable as a per-block indicator for most of the Group C1 population, consistent with the identification difficulties above. Within this general unreliability, Mode 1 behaved more consistently than Mode 0 across specimens – a useful, if modest, positive finding, since it suggests the identification difficulty is not uniform across the two tracked modes and that a future repeat of this campaign might improve reliability simply by prioritising whichever mode proves more identifiable early on, rather than treating all tracked modes as equally trustworthy by default.

Damping ratio, in principle one of the more sensitive indicators to early damage under the CDM ordering (Section 6.1.2), is the indicator most damaged by this unreliability. Damping estimation is more sensitive to noise and to identification error than frequency estimation even under good conditions, since damping is extracted from the width or decay envelope of a resonance rather than its peak location, and both quantities are corrupted by exactly the same identification-quality problems as frequency but with a smaller signal-to-noise margin to begin with. Combined with the fact that damping was not computed for some blocks at all (Section 6.1.1), the resulting record cannot distinguish a genuine damping change from noise, identification error, or environmental drift for the great majority of the campaign, and damping is therefore assessed here as practically unusable as a damage indicator for this dataset, independent of its CDM-motivated theoretical promise. Coherence and peak FRF amplitude, both likewise extracted at the identified resonance frequency, inherit the same dependence and are assessed the same way: broadly consistent with expectations where the identification was clean, but not separable from identification noise where it was not, and therefore not usable as standalone evidence.

By contrast, HDI, amplitude RMS, and force RMS are all insulated from the specific failure mode identified in Section 6.1.1 – ambiguous pole separation under noise and closely-spaced, highly-damped modes – because none of the three requires a modal parameter (a pole, a damping ratio, a mode shape) to be extracted at all. HDI is computed directly from the harmonic content of the response spectrum at multiples of the drive frequency; amplitude RMS and force RMS are computed directly from the raw response and excitation signals respectively. Amplitude RMS and force RMS show a broadly monotonic increase across the run for most specimens, without the block-to-block reversals that characterise the modal-identification-dependent indicators (Section 5.2.4 of Chapter 5).

HDI behaves differently from the other two, and its pooled, all-specimens plot (Figure 5.6) is misleading on its own: because a small number of specimens' final-block values are very large, the plot compresses every other specimen's record toward the bottom of the range, which reads as "flat with two outliers" rather than revealing the pattern that is actually present case by case. Examined per specimen (Figure 6.3), HDI holds at a low, comparatively constant plateau for nearly the entire recorded life, with a rise appearing only in the final block or two before fracture, in every one of the three specimens examined individually (S02, S04, S08). This is consistent with the CDM framework's own account of fatigue life: crack propagation, the mechanism HDI's breathing-crack nonlinearity is intended to track, occupies a much smaller fraction of total life than crack initiation in the high-cycle regime this campaign operates in (Section 2.1 of Chapter 2), so a signal confined to the final block or two before fracture is itself physically plausible rather than a sign that HDI is a poor indicator.

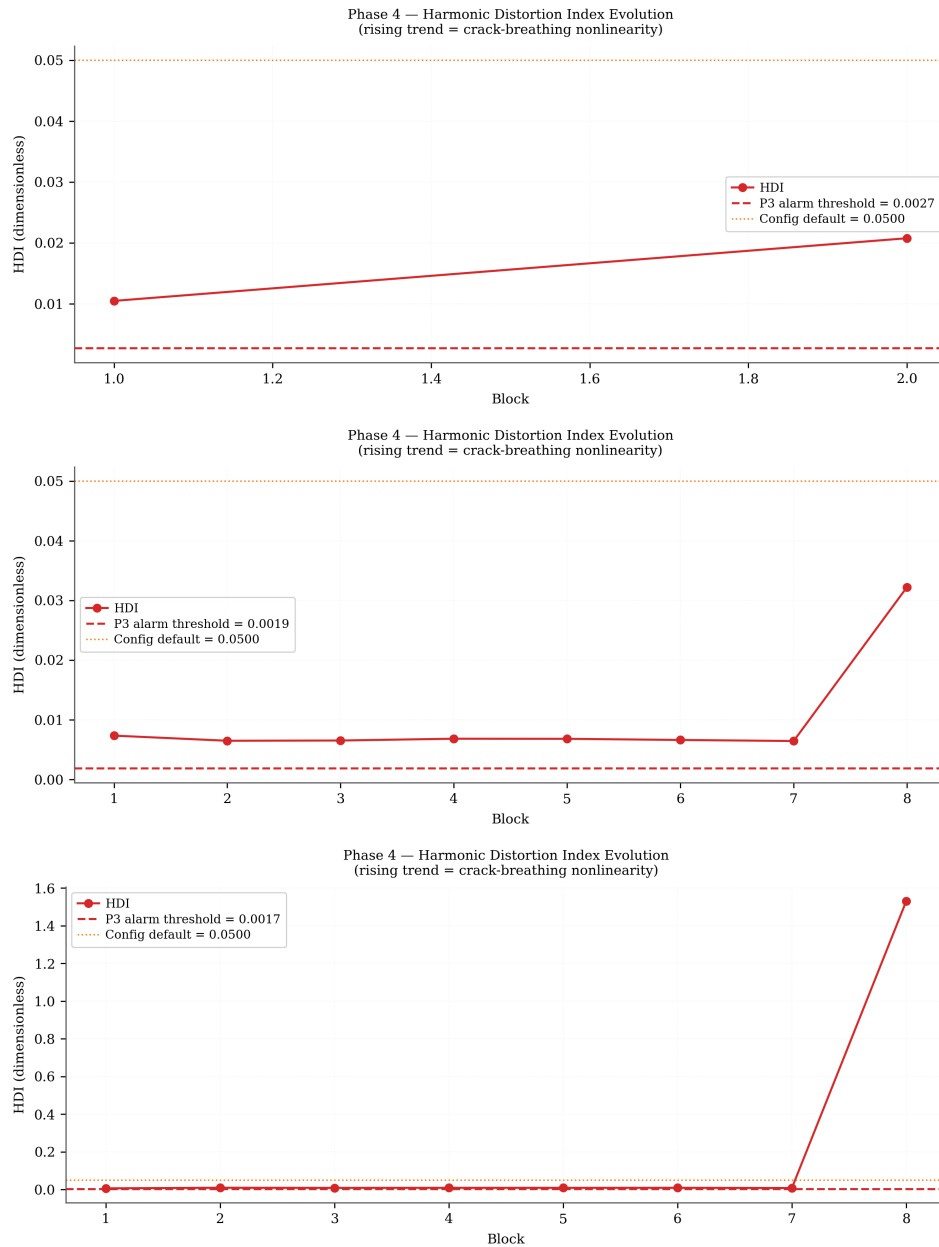


Figure 6.3: Per-specimen HDI evolution through Phase 4 for three representative specimens (S02, S04, S08), each shown on its own scale. All three hold at a low, near-constant plateau for most of their recorded life, with a rise confined to the final block or two before fracture; the magnitude of that final rise differs by more than two orders of magnitude between S02/S04 and S08.

The magnitude of that final-block rise, however, varies by more than two orders of magnitude across the three specimens shown – from roughly a twofold increase for S02 to over $100\times$ the plateau value for S08 – and this variation is not itself attributed to crack propagation alone. For at least S08 (and plausibly S06, which shows a comparably large final-block HDI value in the pooled record), the two fractured halves of the specimen are understood to have made brief post-fracture contact, exceeding the LDV's measurement bounds and inflating the recorded HDI value beyond what the crack-breathing nonlinearity alone would produce. The smaller, more moderate rises shown by S02 and S04 are not affected by this and are read as the cleaner examples of the underlying pre-fracture signal. This distinction matters for how HDI's practical value is assessed: as a fine-grained prognostic indicator, a signal confined to the last block or two before fracture offers little advance warning, and where the largest recorded values are partly a post-fracture artefact rather than a pre-fracture measurement, HDI is better read as a failure-confirmation indicator than as an early or mid-life prognostic one.

Of the three identification-independent indicators, amplitude RMS gives the clearest, most consistently monotonic signal across the recorded life; force RMS behaves similarly but is noisier. HDI's value is concentrated at the very end of life and is partly confounded by the post-fracture artefact just described, so it is not read as comparably prognostic to the other two, though it remains unaffected by the modal-identification problems that degrade frequency, damping, coherence, and peak amplitude. This distinction is itself informative for R1: it indicates that, under the boundary-condition and load-transfer problems documented above, a monitoring approach relying on modal-parameter extraction is considerably more fragile than one relying on directly-measured response and excitation quantities, a distinction the CDM framework's frequency-centric formulation does not itself anticipate – though not every identification-independent indicator is equally useful as a result, since HDI's timing limits its prognostic value even where its computation is robust.

Whether a hardening-then-softening frequency plateau (Lorenzino and Navarro's reported early-life frequency increase, Section 6.1.2) is present in Mode 1's record for any specimen cannot be safely determined from this data: the frequency noise documented throughout this section is severe enough that a genuine early-life plateau would not be reliably distinguishable from ordinary block-to-block scatter, and no such determination is made here.

6.1.4. Phase Indicators: A Mixed Picture

The two phase-based point-tier indicators behaved very differently from one another. Phase quadrature deviation was too noisy to support further analysis, consistent with its status as a second-derivative-like quantity that amplifies whatever noise is present in the underlying phase measurement. Phase-diff slope, by contrast, was well-behaved: close to zero and stable for almost the entire run in nearly every specimen, with the single exception of Specimen 7, whose accelerometer disconnection partway through testing (Section 5.2.4 of Chapter 5) provides a ready, non-damage-related explanation for its divergent record and should be excluded from any population-level statement about this indicator. Where phase-diff slope did depart from its otherwise flat baseline, the departure was modest in magnitude and concentrated toward the end of specimen life, consistent with a genuine but late-onset signal rather than an early one. Spatial phase standard deviation showed a similar, and possibly even cleaner, pattern under the same Specimen 7 exclusion, suggesting that the spatial aggregation this indicator performs may average out some of the noise that limits the point-tier phase measures individually.

6.1.5. Spatial Indicators: Localisation and Mode Shape Distortion

Among the spatial-tier indicators, MAC and modal curvature broadly matched expectations, with some specimen-level outliers, while the localised curvature feature was considerably less well-behaved, showing large, unstructured jumps across the run rather than a consistent trend. Part of the explanation is geometric: the scan grid begins 10 mm beyond the notch and does not span it (Section 3.3.2 of Chapter 3), so the localised curvature feature reports where curvature is greatest *within the measured window* rather than at the crack, and a small change in the deflected shape can move that location from one grid point to another without any corresponding change in damage. This unstructured behaviour is plausibly compounded by the scan grid itself: the nine-point grid used throughout this campaign falls below the 20–30 points identified in the pipeline documentation as the density needed for reliable curvature estimation via central finite differences (Section 3.3.2 of Chapter 3), so the localised curvature feature specifically is operating with less spatial resolution than the method was designed to support, independent of any genuine damage signal; MSE-TVD's spatial averaging across the full scan length makes it comparatively less exposed to this same limitation, which is consistent with its cleaner, more monotonic behaviour relative to the localised curvature feature. MSE-TVD showed a qualitatively distinctive pattern not anticipated by a simple monotonic-damage argument: it rises then falls during Phase 3 (the healthy baseline), reaches a minimum at block 6 – the first Phase 4 block for every specimen, by construction (Section 5.1 of Chapter 5) – and then broadly increases through the remainder of Phase 4. The coincidence of the MSE-TVD minimum with the Phase 3/Phase 4 boundary is unlikely to be a fatigue effect, since Phase 3 involves no net stiffness loss by design; it is more plausibly an artefact of the change in loading condition itself (the step up in drive amplitude between the two phases, Section 3.4.3 of Chapter 3) than of accumulating damage, though this is noted as a plausible explanation rather than a demonstrated one.

The mode-identity problem established in Section 6.1.1 bears directly on these spatial results. Because the MAC and curvature indicators are computed per mode, and because the mode labels do not reliably track the same physical mode between blocks, a discontinuity in a spatial trajectory cannot be distinguished from a relabelling event without independent confirmation that the same mode is being compared. The spatial indicators' strong showing in the sensitivity ranking should be read with that qualification attached.

6.1.6. Comparison with Finite Element Predictions

The Synthetic Data Toolkit's CDM–Basquin prediction (Section 5.4.2 of Chapter 5) provides an independent, model-based trajectory against which the experimental record can be compared, and the comparison is informative precisely because the two disagree. The experimental curves (Figure 5.17 of Chapter 5) are considerably messier than the toolkit's smooth prediction (Figure 5.18 of Chapter 5), for the data-quality reasons already discussed in Section 6.1.1, but the disagreement is also qualitative, not just a matter of scatter: the toolkit's predicted damage trajectory departs from baseline early and then grows at a progressively slower rate (Section 5.4.2), whereas the experimental record stays comparatively flat through much of life before accelerating toward failure (Section 5.4.1) – an early-departing, then-slowng model trajectory against a late-accelerating experimental one. Four candidate sources for this qualitative mismatch are identified, in roughly descending order of expected contribution, and the FEMAP stress field now available for direct inspection (Figure 5.22 of Chapter 5) sharpens the first and third of these from a generic concern into a visually grounded one.

First, the FE mesh used for the toolkit's structural model is coarse, which is expected to smear and advance the apparent onset of stiffness loss relative to a mesh fine enough to resolve the true, highly localised stress and strain gradients at the notch root; Figure 5.22 shows the stress concentration at the notch root spread across a visibly limited number of elements, consistent with a mesh density that would struggle to resolve a sharp, localised gradient at exactly the location where the toolkit's damage kinematics are most sensitive to it. Second, the toolkit's assumed modal damping is likely inadequate for the physical specimens, since damping was set from generic literature values rather than calibrated against this campaign's own (admittedly noisy) damping record. Third, the FE model's notch geometry does not match the final specimen design used for Group C1 (Section 3.2.2 of Chapter 3), a straightforward and avoidable source of discrepancy visible directly in Figure 5.22's notch geometry and one that should be corrected before any future quantitative use of this comparison. Fourth, and most fundamentally, the toolkit's governing assumption is linear-elastic material behaviour, which by construction excludes any plastic deformation or nonlinear load-response effects – exactly the mechanisms most active near a notch root under high-cycle fatigue loading, and plausibly the largest single reason the model's damage trajectory shape differs qualitatively from the measured one rather than merely being noisier; the exaggerated-displacement deformation shown in Figure 5.22 is itself a linear extrapolation of a small-strain solution and would not capture any such nonlinearity even if the mesh and damping issues above were resolved.

Taken together, Figure 5.22 is best read as illustrating *why* the toolkit's prediction diverges from the experimental record in the specific way it does, rather than as independent confirmation that the toolkit is wrong: it is a single static linear-elastic solve under one load case, standing in for the much more complex, cyclic, and locally nonlinear stress state the physical notch root actually experiences over the course of fatigue life. Used this way – as a qualitative diagnostic of the model's own assumptions rather than as a validated stress field in its own right – it is a genuinely useful check on where the toolkit's simplifications are most likely to matter, even though it cannot, on its own, quantify how much of the model-experiment gap each of the four candidate sources above contributes.

A related, separate observation is that the specimens' measured natural frequencies were generally higher than the FE model predicted at baseline (Section 5.6 of Chapter 5). Beyond the notch-geometry and boundary-condition differences already noted, this is also consistent with manufacturing-induced plasticity: cold work introduced during specimen fabrication (cutting, notching) can locally increase yield strength and effective stiffness relative to the bulk material properties assumed in the FE model, which would raise the measured natural frequency above the FE prediction independent of any modelling error in the mesh or damping. This is offered as a plausible contributing explanation, not a demonstrated one, alongside the setup-level differences already discussed.

6.1.7. Diagnostic Evidence from Specimen Groups D1 and D2

The two diagnostic groups described in Section 5.8 of Chapter 5 were produced specifically to test the attribution made in Section 6.1.1: that the clamping arrangement and the load transfer path, rather than the specimens, were responsible for the measurement behaviour of the main campaign. Group D1 established a revised configuration and Group D2 was run under it.

Two conclusions can be drawn, and a third cannot.

First, the Group D2 specimens agree with one another in a way the Group C1 specimens do not. All four share the same FRF shape, with the same resonances at the same frequencies in the same order and the same relative magnitudes (Section 5.8.2). The Group C1 records differ from one another in shape rather than by an offset, and that is what makes them hard to work with: a population-level analysis needs a common baseline against which individual departures can be read as damage, and Group C1 does not supply one. This is the specific sense in which the revised configuration produced a better dataset. It is not that the D2 records are quieter, but that they are comparable.

Second, and following from the first, progression within a specimen becomes legible once cross-specimen agreement is established. D2-03's fatigue run sits approximately 15 Hz below its characterisation runs at the first mode with the mode shape preserved, which is what stiffness loss produces; D2-05's numbered and lettered runs separate in order rather than crossing. Neither observation would be available in a record where the runs disagreed in shape to begin with. The relationship runs one way: agreement is the precondition for reading progression, which is why the Group C1 campaign could establish the second only weakly.

A qualification attaches to both. The largest source of remaining variation within Group D2 is not the fixture, which was identical across all four specimens, but the specimens themselves: the two produced when the usual machining equipment was unavailable show local coherence failures that the other two do not. Specimen manufacturing consistency is therefore a control on measurement quality alongside the fixture, and is taken up as a design recommendation in Section 7.3.5. It is a qualification rather than a counterexample, since the shape agreement and the run ordering survive in those two specimens as well; what degrades is the local quality of the estimate, not the identity of the response.

What cannot be concluded is that the revised boundary condition and load transfer arrangement *caused* the improvement. The D specimens differ from the Group C1 specimens in beam length, drive frequency, test duration and manufacturing batch as well as in fixture, and a shorter beam is stiffer, so a given clamp compliance forms a smaller fraction of the total compliance regardless of any change to the clamp itself. No Group C1 specimen was re-tested under the revised fixture at the original geometry, which is the experiment that would isolate the fixture contribution. The D-group evidence is consistent with the Section 6.1.1 interpretation and would have counted against it had the revised configuration produced equally disordered records; it does not establish it.

6.1.8. What R1 Can Still Be Said to Show

Despite the data-quality limitations documented throughout this section, R1 is not left without an answer. The clearest and most defensible finding is a negative-but-informative one: under boundary conditions and load-transfer quality of the kind encountered in this campaign, indicators that depend on modal parameter extraction (frequency, damping, coherence, peak amplitude) are substantially degraded as damage indicators, to the point of practical unusability for damping specifically, while indicators computed directly from the raw response and excitation signals (HDI, amplitude RMS, force RMS) remain usable and, for amplitude RMS and force RMS specifically, show a sensible, broadly monotonic progression consistent with accumulating damage; HDI is likewise unaffected by the modal-identification problems but is concentrated late in life and only weakly prognostic as a result (Section 6.1.3). This is itself an answer to R1's underlying question about how the vibration response evolves under resonance fatigue: unreliably, and unevenly across indicator types, in a way that tracks each indicator's dependence on clean modal identification rather than any property of the fatigue process itself. A second, more qualified finding is that the two tracked modes were not equally reliable, with Mode 1 behaving more consistently than Mode 0 – a specimen-level, rather than population-level, observation, but one with practical value for how a repeat campaign might allocate its identification effort. Third, the spatial indicators, MAC and curvature in particular, behaved close to expectation for most specimens, sug-

gesting that where mode shape information can be assembled at all, it remains a comparatively robust source of damage evidence even when the corresponding scalar frequency record is not. Finally, the qualitative disagreement with the FE-based prediction is informative in its own right: it identifies mesh resolution, damping calibration, notch geometry, and the elastic-only assumption as concrete, addressable targets for improving the toolkit’s predictive value, rather than leaving the toolkit-experiment gap as an unexplained discrepancy.

6.2. Indicator Sensitivity and Ranking (R2)

Research Question R2 asked which vibration-based indicators are most sensitive to early fatigue damage. The ranked indicator table (Table 5.3) and the Spearman and SHAP rankings (Figures 5.10, 5.11, 5.12) provide the empirical answer. This section interprets that ranking in three steps: what the SHAP and Spearman measures individually say and why they disagree where they do, what the PC1 loading structure reveals about the feature space’s redundancy, and how the hand-engineered ranking compares against the Layer 0.5 autoencoder-based indicators.

6.2.1. Physical Interpretation of the Ranking

By SHAP importance, HDI and curvature (both modes) are the two leading indicators, particularly in the early-life window ($\tau < 0.3$), with MSE-TVD appearing in the global top 5 just ahead of peak FRF amplitude; the phase-family indicators cluster at broadly similar, mid-table importance; and natural frequency and damping sit at the bottom of the ranking in both the global and early-life views. This ordering is consistent with, rather than independent of, Section 6.1’s findings: HDI does not depend on modal parameter extraction and was already identified there as one of the cleanest indicators in this dataset, while frequency and damping are exactly the two indicators shown to be most degraded by the identification difficulties documented in Section 6.1.1. The SHAP ranking is, in this sense, an independent confirmation of the modal-identification-dependence argument made in Section 6.1, arrived at from a purely data-driven feature-importance procedure rather than from the physical reasoning applied there.

The ranking is also stable in composition, though not in scale, between the early-life and global SHAP views: the same top-5 features remain the top-5 in both cases, and HDI’s importance is essentially unchanged (approximately 0.10 in both). The other top-5 features, by contrast, decline somewhat from the early-life view (approximately 0.055–0.080) to the global view (values closer to 0.06), narrowing the gap between the top tier and the mid-table phase cluster (which sits at roughly 0.015–0.020 in both views, against a bottom tier around 0.005). The interpretation offered here is that HDI’s informativeness is sustained throughout life, while the other early top-ranked features (curvature, MSE-TVD) are somewhat more concentrated in their early-life signal and contribute proportionally less once the full life record, including its noisier later portions, is included.

The Spearman ranking, by contrast, is dominated by `mac_mode1` despite that same feature sitting near the bottom of the SHAP ranking, followed by maximum curvature and MSE-TVD; spatial-tier features generally rank above point-tier features; and, consistent with Section 6.1’s finding that Mode 1 was the more reliable of the two tracked modes, Mode 1 features generally outrank their Mode 0 counterparts. Two things stand out as apparent anomalies and are addressed directly, since understanding them is central to answering what these two rankings actually measure. First, HDI does not appear near the top of the Spearman ranking at all, despite dominating SHAP. Second, amplitude RMS and force RMS – shown in Section 6.1 to be among the cleanest, most monotonic indicators in the raw record – rank well below the spatial cluster (amplitude RMS below `fn_mode0` and `coh_mode1`; force RMS near the very bottom).

Both anomalies have the same underlying explanation, and it is a methodological one rather than a sign that the ranking is unreliable. Spearman’s ρ_s measures the strength of a *monotonic* relationship between a feature and τ ; the gradient-boosted tree behind the SHAP values can instead exploit any relationship a decision tree can split on, monotonic or not, including threshold effects, plateaus, and non-monotonic swings. If HDI’s raw trajectory is not cleanly monotonic across its range (plausible given Section 6.1’s account of sudden jumps and identification-related noise across the population), a tree-based measure can still extract strong split-based information from it that a rank-correlation coefficient cannot see. Conversely, Spearman rewards *any* feature with a consistent monotonic drift against τ ,

regardless of whether that drift is caused by accumulating fatigue damage or by something else entirely – rig settling, thermal effects, or other slow, cumulative but non-damage-specific processes. Spatial features derived from mode shape comparison are, by construction, sensitive to exactly this kind of slow cumulative drift in addition to genuine localised damage, which is a plausible reason they dominate a purely monotonic-correlation ranking without necessarily being the most diagnostically meaningful features for damage *detection* specifically. This is offered as the more likely explanation for amplitude RMS and force RMS's low Spearman rank, rather than a claim that these two indicators are poor damage indicators: Section 6.1 already established them as clean and monotonic in the raw data, so a low Spearman rank here says more about what Spearman rewards than about what these two features can detect.

Taken together, SHAP and Spearman are not offering two noisy estimates of the same underlying importance ranking; they are answering two different questions. Spearman asks which features drift most consistently and monotonically with life fraction, which conflates genuine damage sensitivity with any other slow monotonic process a feature happens to track. SHAP asks which features a flexible, nonlinear model finds most useful for predicting τ , which can capture threshold or non-monotonic damage signatures that Spearman cannot, but at the cost of being harder to interpret physically feature-by-feature. Read this way, HDI's SHAP dominance despite its Spearman absence, and the spatial features' Spearman dominance despite mixed SHAP standing, are not contradictions to be resolved in favour of one ranking over the other, but a demonstration that indicator ranking is itself a choice of what kind of sensitivity is being asked for.

6.2.2. Redundancy Structure and Feature Space Geometry

The PC1 loading column of Table 5.3 shows a sharply concentrated structure rather than a broadly distributed one: ΔHDI and HDI itself carry by far the largest loadings (0.986 and 0.164 respectively), ζ_{model} and phase quadrature deviation carry small but nonzero loadings (0.003 and -0.002), and every other feature loads at practically zero. This is consistent with, rather than an independent finding from, the trajectory evidence already discussed: PCA's first component is the direction of maximum variance across the pooled, normalised feature matrix, and a feature with a large-magnitude, consistent trend (HDI, per Section 6.1) concentrates variance along a single direction, while noisy, identification-dependent features scatter their variance across many directions and therefore fail to load heavily onto any one component. ΔHDI 's near-total dominance of PC1 is, in this light, close to a restatement of the finding that HDI is the cleanest large-amplitude signal in the dataset, seen from a different angle: PCA is behaving as expected given the input it was given, rather than revealing an unexpected redundancy structure.

6.2.3. Hand-Engineered vs. Autoencoder-Based Indicators

The comparison between the hand-engineered Layer 0 feature set and the two Layer 0.5 autoencoder health indicators (HI1, HI2) in Figures 5.13–5.14 is the first systematic comparison of learned and hand-engineered vibration-based damage indicators on metallic fatigue specimens under controlled resonance loading, a direct contribution to the literature gap identified in Section 2.8.4 of Chapter 2. The headline result is unambiguous: the LOO-CV validation AUROC for both HI1 and HI2 (0.469 and 0.422 respectively, Section 5.3 of Chapter 5) is at or below the chance level of 0.50. On the standard reading of an AUROC statistic, the CAE-based detector performs no better, and by one measure slightly worse, than random guessing when generalising to an unseen specimen. This result held constant across the bottleneck configurations tested (Section 5.3), which is itself informative: if the poor performance were a methodological problem specific to one architectural choice, varying the bottleneck dimension would be expected to change the outcome materially. That it did not is evidence pointing toward a data-quality explanation rather than an architecture-selection one, consistent with the broader data-quality picture established in Section 6.1.1.

Two further observations from the raw HI1/HI2 trajectories support this reading, and both benefit from being made explicit about what the two indicators' values actually mean, since neither is normalised to an interpretable $[0, 1]$ scale. HI1 is a raw resonance-weighted reconstruction MSE, ranging roughly 0.2–1.5 across the population against a Phase 3-derived threshold of 1.359; its absolute scale depends on signal preprocessing choices with no inherent physical meaning of its own, which is exactly why the

pipeline defines its alarm criterion as a threshold-crossing (the 99th percentile of the Phase 3 baseline) rather than as an absolute value. HI2 is explicitly a maximum z -score across the latent dimensions, ranging roughly 2.25–4 against a threshold of 3.529; because it is already a standardised quantity, its non-trivial baseline value is not itself an anomaly, but the specific baseline level is informative. Taking the maximum over many latent dimensions is an extreme-value statistic, and the expected maximum of a collection of approximately standard-normal variables grows with the number of dimensions being maximised over; HI2 sitting at 2.25 or above at baseline, rather than near zero, is therefore partly a structural consequence of maximising over a 64-dimensional latent space, not solely a symptom of noisy input data. This also offers a direct explanation for why performance did not improve materially at smaller bottleneck sizes: a smaller latent space reduces the number of dimensions being maximised over, which would mechanically lower the extreme-value baseline, but does not by itself make the underlying reconstruction or latent deviation any more diagnostic of genuine damage, so the detection criterion does not become reliably easier to satisfy just by changing dimensionality. Both HI1 and HI2 remain below their respective p99 thresholds for most of most specimens' runs, consistent with a baseline that is itself noisy enough to set an unusually high threshold – directly supporting the near-chance AUROC finding above, since a threshold calibrated against a noisy baseline will be a poor discriminator regardless of how genuinely informative the underlying reconstruction error is.

The two indicators' trajectories also differ from each other in character, not just in absolute scale. HI1 rises broadly monotonically over the first four blocks before levelling off for most specimens, a pattern broadly consistent with (if noisier than) the CDM-motivated trajectory shape expected of a genuine damage indicator. One specimen shows a distinct peak at block 5 that coincides with a day-to-day testing break: this specimen was tested across two separate sessions (Section 5.1 of Chapter 5), and the first block of the second session appears to trigger an elevated HI1 reading, consistent with a transient artefact of the pause and restart (re-mounting, thermal re-equilibration) rather than a genuine damage event, since no comparable jump is expected from the fatigue process itself at that particular life fraction. HI2, by contrast, shows a considerably messier record, with no consistent shape across specimens: some rise, some fall, and some remain roughly constant through the run. This asymmetry suggests that the reconstruction-error perspective (HI1) retains at least a qualitatively plausible trend shape even where its absolute discriminative power is poor, while the latent-deviation perspective (HI2) does not even preserve a consistent trend shape across the population – a further piece of evidence, independent of the LOO-CV AUROC result, that the CAE's learned representation is not capturing a population-consistent damage signature in this dataset.

Hand-Engineered vs. Autoencoder: A Comparative Synthesis

Read against the broader literature (Section 2.8.4 of Chapter 2), which motivates autoencoder-based indicators specifically for their expected advantage on non-Gaussian or nonlinear departures from a healthy manifold that linear, hand-engineered features may miss, this campaign's result is a clear instance of that theoretical advantage failing to materialise in practice, and the reason is traceable rather than mysterious. The hand-engineered indicators that performed well in this campaign (HDI, curvature, MSE-TVD) succeeded specifically because they do not depend on clean modal parameter extraction (Section 6.1); they are relatively simple, targeted statistics computed directly from the measured signal or mode shape. The autoencoder, by contrast, is asked to learn a general-purpose representation of the entire healthy-state FRF or feature vector from a small population (eight specimens) with a noisy baseline, and its failure here is consistent with exactly the scenario the CDM/modal-identification-dependence discussion in Section 6.1 would predict for any data-hungry, general-purpose learned method applied to a small, noisy population: without enough clean specimens to characterise what "healthy" looks like precisely, a learned reconstruction target is itself built on unreliable ground, and no amount of architectural tuning (bottleneck size, in this case) can substitute for that. The broader methodological lesson is not that autoencoder-based indicators are unsuited to this problem in general, but that they are considerably more data-hungry and more sensitive to baseline data quality than the targeted, physically-motivated hand-engineered indicators that worked best here – a genuinely useful, if negative, empirical finding for anyone considering this class of method for a similarly small, noisy fatigue-monitoring campaign.

6.3. Damage Index Behaviour and CDM Model Consistency

The six damage index trajectories in Section 5.4 show how different representations of the same underlying damage process agree and diverge, and this comparison is best read as a continuation of Sections 6.1–6.2 rather than as an independent result: since every DI formulation is built from the same feature set already shown there to be of mixed and generally poor reliability, the DI-level findings largely inherit, rather than newly establish, those data-quality conclusions.

6.3.1. Which Formulation Best Represents Damage Progression?

One constraint applies before any formulation is compared against another. Every one of the six converges for exactly five of the eight specimens, and the three that fail are the same three in each case: S01–S03, which completed only one or two Phase 4 blocks before fracture and therefore offer too few points to fit a two-parameter logistic at all (Section 5.4.1 of Chapter 5). The comparison below is therefore between formulations evaluated on five specimens, and no formulation could have distinguished itself on the other three. Among the fits that do converge, the recurrence of exact parameter bounds discussed in that same section means a subset describe the logistic form failing to find a sigmoid rather than a measured transition. Both facts narrow what “best represents damage progression” can mean here: it is a judgement about behaviour on a five-specimen subset, part of which rests on degenerate fits.

The three formulations that Section 5.4.1 identifies as producing the most consistently S-shaped trajectories, D_{comp} , D_{PC1} and D_{CAE} , share a property that explains why. Each case can be traced to a specific formula. $D_{\text{PC1}}(\tau)$ is the projection of the full feature vector onto the PC1 loading direction $\mathbf{v}_1$ (Section 4.3.2 of Chapter 4); Section 6.2.2 already established that this direction is overwhelmingly dominated by ΔHDI and HDI (loadings 0.986 and 0.164, essentially all other features near zero), so D_{PC1} is, in practice, close to a rescaled HDI trajectory rather than a genuinely multi-feature aggregate. $D_{\text{comp}}(\tau)$ is an explicit SHAP-importance-weighted sum over the top- k non-redundant features, with weights $w_j = \Phi_j^{\text{global}} / \sum_i \Phi_i^{\text{global}}$ (Section 4.3.4 of Chapter 4); because HDI and curvature dominate the global SHAP ranking (Section 6.2.1), D_{comp} is, by construction, weighted most heavily toward exactly the indicators already identified as the cleanest in this dataset. D_{CAE} (reconstruction) is not built from named features at all, but its per-specimen trend consistency is plausibly connected to the same underlying pattern via a different route: Section 6.2.3 found that HI1 specifically (not HI2) retained a qualitatively plausible, broadly monotonic trend shape across most specimens despite the CAE’s poor cross-specimen generalisation (LOO-CV AUROC at or below chance). D_{CAE} ’s within-specimen consistency is therefore consistent with that HI1-specific finding, not a separate or contradictory result.

The general conclusion is that all three formulations appear consistent *because* they are constructed to track, or happen to correlate with, the small subset of features already shown in Sections 6.1–6.2 to be robust to this campaign’s data-quality problems. Their apparent quality is therefore better read as a measure of the underlying metric’s robustness to noise than as direct evidence of superior damage-detection capability: a formulation dominated by one clean, large-amplitude, but not exclusively damage-specific signal (HDI) will look consistent regardless of whether HDI itself is the most diagnostically meaningful indicator of the underlying fatigue process, a distinction already raised for exactly this reason in Section 6.2.1.

This same distinction bears on the prognostic horizon result taken up in Section 6.4.1. “Consistent” as used here is a within-specimen property, meaning a clean monotonic per-specimen trajectory, whereas reaching a population τ_h requires a cross-specimen property: enough specimens crossing the CI-narrowing criterion at a comparable life fraction for the population Gaussian Process to resolve a posterior. A formulation can be clean for each specimen individually while varying widely between specimens in *when* that cleanness sets in, which is what the boundary-condition and load-path variability of Section 6.1 would be expected to produce. The horizon result itself, and the reasons none of the six formulations reached one, are discussed in Section 6.4.1 rather than here.

6.3.2. Comparison with the Synthetic Data Toolkit Prediction

The comparison between the experimental $D_{\text{freq}}(\tau)$ record and the toolkit prediction reproduces the mismatch already analysed for natural frequency in Section 6.1.6, and for the same four reasons given

there: the toolkit’s predicted damage index is approximately logarithmic in cycle count, rising quickly and then flattening, whereas the experimental record stays comparatively flat before accelerating toward failure. It is not independent evidence, since both comparisons draw on the same underlying model and the same physical record.

One qualification must now be attached to this comparison that was not previously apparent. The experimental side of it is the $\hat{\tau}_{\text{mid}}$ obtained from the logistic fits of Table 5.4, and Section 5.4.1 establishes that a substantial fraction of those fits terminated at a parameter bound rather than at an interior optimum, with $\hat{\tau}_{\text{mid}}$ values of exactly 0.000 and 1.000 recurring across formulations. Where that has happened the fitted inflection point records the logistic form failing to find a sigmoid in the data, not a measured damage transition. The direction of the model–experiment disagreement is not in doubt, since it is visible in the raw trajectories, but its magnitude should not be quoted from the fitted parameters.

6.4. Remaining Useful Life Estimation (R3)

Research Question R3 asked whether vibration response trends can be used to estimate remaining fatigue life. Because RUL estimation in this pipeline is built directly on the same DI formulations discussed in Section 6.3, the data-quality conclusions reached there carry forward largely unchanged: the RUL results are broadly consistent with, rather than independent of, the modal-identification-dependence findings of Sections 6.1–6.3, and the overall confidence achievable in any individual RUL prediction remained low throughout this campaign.

6.4.1. Prognostic Horizon: What It Means in Practice

None of the six DI formulations reached a resolved population-level prognostic horizon within observed specimen life (Section 5.5 of Chapter 5), D_{freq} included despite its direct tie to the CDM stiffness-loss relationship (Section 6.3.1 above). Two mechanical points from the prognostic horizon definition itself help explain why this held uniformly rather than only for the less internally consistent formulations. The horizon condition is $\Delta_{\text{CI}}(\tau) < 0.30$ (Section 4.4 of Chapter 4), the width of the 95% confidence interval on the RUL estimate, and this width can only narrow below threshold if the population GP has enough consistent cross-specimen structure at a given τ to constrain its posterior.

Before that explanation is accepted, however, a more mundane one must be set alongside it. For three of the six formulations the reported confidence-interval widths are not merely wider than the 0.30 threshold but numerically meaningless, reaching magnitudes many orders larger than the total life they estimate (Section 5.5.1 of Chapter 5). This is a failure of the linearised delta-method propagation of Section 4.4.4 on a near-singular Jacobian, not a statement about damage. For those formulations the horizon criterion could not have been satisfied at any block however consistent the underlying trajectories were, so their failure to resolve a horizon carries no information about detectability, and the argument below applies only to the formulations whose uncertainty estimates remained finite. For a repeat campaign, fixing the uncertainty propagation precedes any question about indicator choice.

A second constraint compounds this: the uncertainty analysis covers only four of the eight specimens, since S01–S03 yielded too few Phase 4 blocks to fit at all. The population-level result therefore rests on half the population, and on the half with the longest lives, which is the half least representative of the loading conditions the protocol was designed to explore. A formulation whose trajectory shape and timing vary idiosyncratically from specimen to specimen – which Section 6.1 argues is expected of every formulation in this campaign, given specimen-to-specimen boundary-condition and load-transfer variability, not just the feature-combination formulations – offers the GP comparatively little of this shared structure to exploit regardless of which underlying indicator is used. D_{freq} ’s population-general *scale* of change (Section 6.3.1) is a necessary condition for the GP to have consistent structure to learn from, but Section 6.4.3 argues it was not, on its own, a sufficient one given this campaign’s population size and per-specimen data quality. This is consistent with, not contrary to, the broader picture established across Sections 6.1–6.3: reliable population-level prognosis in this campaign depended on a degree of cross-specimen timing consistency that none of the six formulations, including the physically best-motivated one, achieved.

6.4.2. Method 1 vs. Method 2: Trend Extrapolation vs. Gaussian Process

The comparison between Method 1 (single-specimen parametric extrapolation) and Method 2 (population Gaussian Process) matched the expected pattern given the population size available. With only eight Group C1 specimens entering the leave-one-out cross-validation, the population Gaussian Process has enough shared structure to produce a usefully narrowed posterior for some specimen and formulation combinations but not reliably across the board, which is consistent with none of the six formulations reaching a population-level prognostic horizon. Table 5.5 of Chapter 5 bears this out: Method 2 outperforms Method 1 at some formulation and checkpoint combinations and not others, rather than either method being consistently superior. That qualitative unevenness, rather than a claim of overall superiority for either method, is the defensible reading given this population's size and noise level. The comparison is in any case not like-for-like, since Method 2's values are cross-validated on a held-out specimen while Method 1's are in-sample errors for the specimen being fitted (Section 5.5.2).

6.4.3. Population Size, Data Quality, and the Practical Reliability of RUL Estimates

One clear, positive pattern emerges despite the population-level and data-quality limitations documented throughout this chapter: RUL estimation performed better for the longer-lived specimens, in the specific sense that a usable, resolved estimate was more often obtainable at all. Two concrete mechanisms explain this, both following directly from how Method 1 and Method 2 operate. First, a longer-lived specimen simply produces more Phase-4 blocks before fracture, giving Method 2's population GP more checkpoints of data to condition on for that specimen and giving Method 1 a longer record to fit its extrapolation against. Second, a longer-lived specimen is, by construction, one whose damage indicators evolved more mildly (a smaller effective rate of change per block) rather than rising sharply toward failure – and a milder, more gradual trajectory offers both methods more informative early-life data points to extrapolate or condition from, whereas a rapid, near-vertical late rise (characteristic of the shorter-lived specimens) leaves comparatively little useful signal before the specimen fractures. Despite this, confidence in any individual RUL prediction remained low throughout the campaign: the population size (eight specimens) is small relative to what the broader GP-based prognostics literature typically validates against, and this campaign's own data-quality problems (Sections 6.1–6.3) compound rather than merely coexist with that small- n limitation, since a small population of noisy trajectories offers the GP even less consistent structure to learn from than a small population of clean ones would. Chapter 7 returns to expanded population size as the primary recommendation for improving this result in any future repeat of this campaign.

6.5. Resonance Testing as a Monitoring Platform (R4)

Research Question R4 asked how resonance-based vibration fatigue testing compares to traditional fatigue tests. This section addresses R4 in two parts: the S-N characterisation of the campaign itself, read against its own methodological limitations before any literature comparison is attempted, and the fixed-frequency protocol's viability as a monitoring platform. The Group B specimens are not treated as a separate comparison point in this section: Section 5.1 of Chapter 5 already established that Group B was run under a heterogeneous, iteratively-tuned protocol rather than one fixed configuration, which is precisely why it was excluded from the Group C1 population statistics; revisiting it here as an S-N cross-check would reopen a comparison already judged not meaningful. Group B data is used once in this section (Section 6.5.2), but only as qualitative mechanistic evidence for the resonant-amplification argument, not as a fatigue-life population to compare against Group C1.

6.5.1. S-N Characterisation: Method Limitations Before Literature Comparison

Before comparing the Basquin fit in Figure 5.23 against the literature, the limitations of the method used to produce it need to be stated plainly, since they bear directly on how much weight the fit itself can carry. Three limitations compound.

First, the force-to-stress calibration factor k (Section 5.6 of Chapter 5) was obtained from a linear-elastic finite element analysis applied to a geometry that is not itself linear in its response: the V-notch introduces a local stress concentration and, at the higher end of the tested amplitude range,

plausibly some local plasticity, neither of which a linear FE stress solve captures. Compounding this, the modal characterisation of the FE model and the physical specimens differed substantially in some cases (Section 6.1.6), which is itself evidence that the FE model's stiffness distribution is not a fully faithful representation of the physical specimen it was meant to calibrate. A further point about these values should be stated rather than left to the table: although the calibration was per-specimen in intent, only three distinct k values resulted across the eight specimens, with six of them sharing a single value (Section 5.6.1 of Chapter 5). For three quarters of the population the conversion from force to stress is therefore a single shared factor, and the specimen-to-specimen variation the calibration was meant to capture is largely absent from it. This is a limitation of the calibration method, not of the specific numerical values obtained: the k values in Table 5.6 are the actual FEMAP output for this campaign, not placeholders, but the linear-elastic assumption underlying them means they should still be read as a first-order calibration rather than a fully validated one.

Second, the assumption that the shaker delivers a pure transverse bending load is itself questionable for this campaign. Because the specimen's response geometry is not linear at larger amplitudes, the component of the applied force actually transmitted to the specimen changes with amplitude: rather than remaining a pure bending excitation, part of the applied force plausibly acted axially, alternately compressing and tensioning the specimen along its length rather than only bending it transversely. This is not merely a theoretical possibility. Section 6.1.1 establishes from the mounting-arm fracture surface (Figure 6.2) that the attachment was carrying a bending moment substantial enough to fail the arm in fatigue, and any moment the arm reacted was also applied to the specimen. The consequence for this section specifically is that the load cell records only the component along the intended axis, so the unrecorded component enters no stage of the stress calculation. If the load applied to the specimens themselves included a comparable axial component, the nominal stress amplitude computed from the measured transverse force alone would misrepresent the true stress state at the notch root, independent of and in addition to the k -calibration uncertainty above.

Third, the campaign's setup consistency and sample size are modest relative to standard fatigue-testing practice. Standard test methods for axial fatigue testing of metallic materials, such as ASTM E466 (ASTM International, 2021) and ISO 1099 (International Organization for Standardization, 2017), specify controlled, uniaxial load application and recommend enough replicate specimens per stress level to characterise scatter with reasonable confidence, typically substantially more than the two to three replicates per nominal force level achieved here (and, per Section 5.6.3 of Chapter 5, the achieved drive amplitudes varied enough within each intended level that the eight specimens spread across five distinct stress levels rather than three, further thinning the replication at any one level). Against this standard, the present campaign's Basquin fit ($R^2 = 0.0227$, Section 5.6) should be read as a weak, exploratory characterisation rather than a validated S-N curve, and the limitations above – provisional stress calibration, a plausible mixed loading state, and thin replication – are jointly sufficient to explain a fit this poor without needing to invoke any additional, unidentified source of scatter.

With this context established, the fatigue lives obtained in this campaign can be compared, cautiously, against the published literature. Yahr's compiled design curve for 6061-T6 aluminum under fully-reversed loading (Yahr, 1997) reports a fatigue strength of approximately 129 MPa at 2×10^5 cycles, falling to an asymptotic high-cycle value in the region of 50 MPa by 10^7 – 10^9 cycles. Read against this reference point, the eight Group C1 specimens (Table 5.6 of Chapter 5) split into two groups, and both groups show the same directional shortfall. The four specimens tested above 129 MPa (C1-01 at 232.08 MPa, C1-05 at 183.21 MPa, C1-06 and C1-07 both at 193.87 MPa) all failed short of 2×10^5 cycles, as expected at stress levels above the reference point – but not by a small margin: their fatigue lives range from 6,930 to 112,500 cycles, all well under half of the reference life even for the specimen closest to it. The four specimens tested at or below 129 MPa (C1-02 at 134.06 MPa, C1-03 at 112.06 MPa, C1-08 at 109.50 MPa, and C1-04 at 83.94 MPa) are the more striking case: at these stress levels, Yahr's curve would predict lives at or beyond 2×10^5 cycles, trending toward very long life as stress approaches the asymptotic 50 MPa region. Instead, all four failed between 11,880 and 59,250 cycles – roughly one to two orders of magnitude short of the literature expectation, with C1-04 (the lowest-stress specimen in the population, and the one furthest below the reference point) showing no corresponding extension of life. The shortfall is therefore not uniform across the population: it is proportionally worse at lower stress, exactly where the method limitations above (an uncalibrated, linear-elastic stress estimate; a possible unaccounted axial load component; thin replication) would be expected to matter most relative

to the true underlying signal. Given the three method limitations above, a lower-than-literature result of this kind is not surprising: an uncalibrated or under-calibrated stress estimate, a load state that may include an unaccounted axial component, and a small, heterogeneously-scattered sample would each independently tend to bias an apparent fatigue life downward or widen its scatter, so this comparison is offered as consistent with, rather than contradicting, the method limitations already identified, rather than as evidence of some additional, unexplained weakness in the material or specimen geometry itself.

6.5.2. The Fixed-Frequency Protocol: Amplification vs. Stress Control

The central trade-off of the fixed-frequency protocol – amplified damage indicator sensitivity in exchange for non-constant stress amplitude – is the core of R4. A fixed drive frequency encodes the resonance detuning directly into the response phase and amplitude, producing a self-amplifying damage signal: as the natural frequency drifts, the phase deviation and response amplitude change automatically without any change to the excitation, whereas a tracked-frequency protocol that continuously re-tunes the drive to remain at resonance would actively suppress this same detuning signal.

The evidence available from this campaign for the amplification argument is more direct, if less complete, than a ranking-based test would provide. On the ranking evidence specifically: phase-family indicators occupied only mid-table SHAP and Spearman importance (Section 6.2), not the top ranks that a strong version of the amplification argument might predict, so this particular line of evidence is inconclusive rather than confirmatory and is reported as such rather than reframed to fit the expected outcome. The stronger evidence is instead direct: the FRF amplitude at resonance remained substantially elevated relative to off-resonance excitation throughout Phase 4 for the great majority of specimens (Section 6.1.8), indicating that dynamic amplification was sustained for the duration of the test rather than lost as the natural frequency drifted. A further, independent piece of evidence comes from the Group B load-configuration tuning data (Section 3.4.3 of Chapter 3), used here strictly as mechanistic evidence rather than as a fatigue-life comparison: specimens excited near the tip showed markedly less damage than specimens excited nearer the mode's antinode, even at comparably large excitation amplitudes. This is direct evidence that the damage rate is governed by the dynamic, mode-shape-weighted response the excitation actually produces, not by the raw excitation amplitude alone – exactly the distinction the resonant-amplification argument for the fixed-frequency protocol depends on.

A direct, controlled comparison between the fixed-frequency protocol used here and a tracked-frequency alternative was not performed in this campaign, and this limits how far the amplification argument can be pushed empirically: everything reported above is evidence that resonant amplification was present and sustained under the fixed-frequency protocol, not a demonstration of how much of the observed indicator sensitivity would be lost under a tracked-frequency protocol specifically. A direct fixed-versus-tracked comparison, ideally on matched specimens from the same population, is identified here as a concrete, well-motivated direction for future work (Chapter 7), rather than a gap this campaign's data can close on its own.

6.6. Fractography Validation: What the Optical Evidence Says

The two Group C2 specimens provided an independent, vibration-free check on the detection timing claims of the Layer 0 and Layer 0.5 pipeline. The optical fractography results in Section 5.7 can now be interpreted against the pipeline's own claims for those specific specimens. As with the results discussed in Sections 6.1–6.5, the outcome here is mixed, but for a different reason: the limitation is one of resolution and timing rather than noise or identification difficulty.

6.6.1. Did the Detection Coincide with Real Damage?

The central question addressed by the Stage 3 microscope images is whether the pipeline's first damage alarm corresponded to a physically observable change at the specimen surface. The answer for both Group C2 specimens is a qualified yes: where the alarm fired, a crack was indeed present at the notch root, but it had already propagated to a visually obvious extent rather than being a subtle, just-initiating feature. This is a genuinely useful, positive result – it rules out the possibility that the alarm

was responding to pure noise or drift, since a real physical change was present at the interruption point in both cases – but it is a more limited result than a confirmation of early sensitivity would be. A third qualification, developed in Section 5.7.2 of Chapter 5, is that the alarm in both specimens was driven by Amplitude RMS rather than by any of the modal indicators ranked highly in Section 6.2: the live monitor operates on the harmonic-loading stage and has no scan-grid data available to it, so the spatial features that dominate the ranking play no part in the detection validated here. What this evidence establishes is that the indicators are consistent with damage; what it does not establish is whether the indicators were also responding, earlier and more subtly, to damage that had not yet become visible at the optical magnification used here. Resolving that second question would require a substantially higher-resolution examination (electron microscopy, at minimum) applied at multiple points earlier in specimen life than the single interruption point this protocol provides, and is identified here as a concrete limitation rather than glossed over.

An informal but informative piece of additional evidence bears on this same question: an event camera being tested on-site by a visiting collaborator during one of the C2 test sessions captured the crack formation directly. The footage shows the crack forming late into the specimen's life and then propagating relatively quickly once formed, rather than growing slowly and visibly over an extended fraction of life. This is consistent with, and helps explain, the "late but real" character of the Stage 3 finding above: if the crack genuinely does not become macroscopically visible until relatively late in life, and then propagates quickly once it does, a protocol that interrupts at the first statistical alarm and inspects once would be expected to find a visible, already-advanced feature rather than catching an earlier, subtler one – not because the indicators lack early sensitivity, but because there may be very little life-fraction window between "not yet visible" and "visible and advancing quickly" for this specimen geometry and loading condition. This observation is offered as a plausible, video-supported explanation for the pattern observed, rather than as a controlled, quantified measurement of crack growth rate.

6.6.2. Fracture Surface Consistency and Protocol Validity

The Stage 4 fracture surface comparison addresses a prerequisite for the Stage 3 result to be interpretable: whether the interruption and restart altered the failure mechanism. For both Group C2 specimens, breaking the samples did not reveal any additional information beyond what Stage 3 had already shown: the final fracture surfaces are qualitatively consistent with the Group C1 reference specimens (Section 5.7 of Chapter 5), with no distinct initiation site, propagation character, or final-fracture appearance (Figures 5.30, C.6) that would suggest the interruption and reload altered the failure mechanism. This supports treating the Group C2 specimens' Stage 3 observations as representative of what an uninterrupted Group C1 specimen would show at a comparable life fraction, rather than as an artefact specific to the interrupted protocol. With only two Group C2 specimens, this remains an indicative rather than a statistically definitive check, and an expanded interrupted-specimen population is identified in Chapter 7 as a direction for strengthening it.

6.6.3. Cross-Specimen Fractographic Evidence for the Fatigue Mechanism

Beyond the C2-specific detection-timing validation above, the fracture surfaces available across the campaign (the Group C1 reference specimens and the two Group C2 specimens, Figures 5.30 and C.6) provide independent, classical fractographic evidence that fatigue was indeed the governing failure mechanism throughout, rather than static overload or a manufacturing defect. Three observations support this. First, specimens with longer fatigue lives show visibly shinier fracture surfaces than specimens that failed early, a qualitative impression visible to the naked eye (most clearly between the Group C1 reference specimens, Section 5.7 of Chapter 5) rather than one quantified with a surface-roughness measurement in this campaign: a smoother, lower-roughness surface is the expected signature of a stable fatigue crack growth zone that has experienced many cycles of crack-face contact and rubbing, with the smoothness increasing, as observed here, the longer that process continues before final fracture. Second, every examined fracture surface shows a clearly demarcated final-fracture region, distinct in character from the surrounding stable-growth zone, consistent with the expected transition from slow, stable fatigue crack propagation to rapid unstable fracture once the remaining ligament can no longer support the applied load. Third, beach marks – the macroscopic arrest-line banding characteristic of progressive fatigue crack growth – are present on every examined specimen. Taken together, these three independent, classical fractographic signatures corroborate the same conclusion

from a source entirely independent of the vibration-based indicators: the dominant failure mechanism throughout this campaign was indeed progressive fatigue crack growth, not an alternative failure mode that might otherwise explain the observed indicator behaviour.

This section's evidence connects directly to one specific finding from earlier in this chapter, worth making explicit rather than leaving implicit: the video-supported observation that cracks formed late and propagated quickly (Section 6.6.1) is consistent with, and offers a physical explanation for, Section 6.4.1's finding that none of the six DI formulations reached a resolved population-level prognostic horizon. A crack that remains invisible until late in life and then propagates rapidly leaves comparatively little life-fraction window in which any indicator – however sensitive – could accumulate the consistent, population-general signal a resolved prognostic horizon requires; this is an independent, physically-grounded reason for the RUL results being generally poor that does not depend on the modal-identification or data-quality arguments made in Sections 6.1–6.3. The broader synthesis of how all of this chapter's findings fit together is left to Section 6.7.

6.7. Importance of Findings and Contribution to the Field

This section situates the R1–R4 results within the three gaps identified in Chapter 2 and the broader claims made in Chapter 1, reading the chapter's findings as a whole rather than section by section. Consistent with the tone established throughout this chapter, each gap is assessed honestly against what was actually found, rather than against what the original research design anticipated finding.

6.7.1. Closing the Three Literature Gaps

Gap 1 (Section 2.10.1) – No systematic multi-indicator comparison under fixed-frequency resonance testing on metals. This gap is closed, and closed more informatively than a comparison under ideal conditions would have been. The ranked indicator table (Table 5.3) and the SHAP/Spearman analysis (Section 6.2) provide the first simultaneous characterisation of point-tier and spatial-tier modal indicators through the complete fatigue life of metallic specimens under fixed-frequency harmonic resonance excitation. What makes this closure substantive rather than nominal is that it does not stop at a ranking: Section 6.1 traces each indicator's behaviour back to a specific, identifiable cause – modal-identification dependence for frequency, damping, coherence, and peak amplitude; direct, identification-independent computation for HDI, amplitude RMS, and force RMS – rather than reporting an unexplained scatter of importance scores. A ranking obtained under favourable laboratory conditions would not have revealed this distinction, since every indicator would have performed comparably well; it is precisely because this campaign's experimental conditions were imperfect that the ranking is diagnostic of *why* indicators differ, not just *that* they differ, which is arguably more transferable to other resonance-fatigue setups than a ranking obtained under conditions few future practitioners will be able to replicate exactly.

Gap 2 (Section 2.10.2) – Absence of a prognostic horizon analysis for modal-parameter-based RUL estimation on metals. This gap is addressed rather than closed. None of the six DI formulations tested reached a resolved population-level prognostic horizon within observed specimen life (Section 6.4.1), D_{freq} included despite its direct tie to the underlying CDM stiffness-loss relationship. This is a considerably more modest outcome than a demonstration of reliable multi-formulation prognosis would represent, and it is presented as such rather than reframed as a broader success. Its value to the field is better understood as a first, quantified account of *how* prognostic reliability degrades under realistic constraints – small population size, imperfect boundary conditions, a crack that forms late and propagates quickly (Section 6.6) – than as a validated methodology ready for direct application. This is still a genuine contribution: no prior study appears to have quantified where and why modal-parameter-based prognosis breaks down for metallic specimens under these conditions, and that quantification is itself useful to a field that has, to this point, mostly reported successes under more controlled conditions.

Gap 3 (Section 2.10.3) – The fixed-frequency protocol as a monitoring platform. This gap is addressed unevenly. The quantitative link between the fixed-frequency protocol and classical S-N characterisation (Section 6.5.1) is weak, constrained by a provisional force-to-stress calibration and a fit with essentially no explanatory power ($R^2 = 0.0227$); this part of the gap remains substantially open. The qualitative evidence for resonant amplification as a monitoring mechanism (Section 6.5.2)

is considerably more solid: the sustained FRF amplitude through Phase 4 and the tip-versus-antinode excitation comparison together demonstrate that dynamic, mode-shape-weighted amplification, not raw excitation amplitude, governs the damage rate, which is a real, if qualitative rather than quantitative, contribution to understanding the fixed-frequency protocol's mechanism.

6.7.2. Scientific Contribution and Broader Significance

Chapter 1 stated that the scientific contribution of this thesis is an integrated, empirically grounded characterisation of modal-parameter-based fatigue prognosis under controlled resonance conditions. The results reported here support the "integrated" and "empirically grounded" parts of that claim more strongly than they support any implicit expectation of a clean, successful outcome, and the contribution is best framed accordingly.

The pipeline is genuinely integrated: detection (Layer 0), sensitivity ranking (Layer 1), damage indexing (Layer 2), and prognosis (Layer 3) connect coherently from one stage to the next, and, importantly, they disagree with each other in traceable rather than arbitrary ways. The indicators found most robust in Layer 0 (Section 6.1) are exactly the ones that dominate the Layer 1 ranking (Section 6.2) and drive the Layer 2 formulations built to concentrate on them (Section 6.3.1); that none of these formulations, nor D_{freq} despite its more direct physical motivation, reaches a Layer 3 population horizon is itself traceable to the same cross-specimen timing variability documented at Layer 0, rather than an unexplained failure appearing only at the prognosis stage. This chain of traceable causation, rather than a chain of uniformly good results, is what "integrated" means in practice for this thesis, and it is a stronger, more falsifiable claim than "the layers agree" would have been.

The results are also genuinely empirically grounded, in that every major finding was checked against at least one independent source rather than accepted on the strength of the pipeline's own internal metrics alone: the FE/synthetic toolkit comparison (Sections 6.1.6, 6.3.2), the optical fractography validation (Section 6.6), and the S-N literature comparison (Section 6.5.1) each independently corroborated or qualified a claim made elsewhere in the chapter, and in no case was an independent check simply omitted because it was inconvenient to the pipeline's own results.

Beyond the specific empirical findings, three further, more tangible outputs of this thesis are worth stating explicitly as contributions in their own right, independent of how the experimental campaign itself turned out:

- First, three separate, reusable software tools were developed over the course of this work: the LDV Analysis Suite (pre-campaign sensor and specimen characterisation), the Synthetic Data Toolkit (the physics-informed CDM–Basquin prediction generator used throughout this chapter as an independent comparison), and the full five-layer analysis pipeline itself (Layers 0–3, live monitoring, and the S-N module). All three are documented in Chapters 3–4 and are available for reuse or extension by future work on this or a similar experimental setup, independent of this specific campaign's results.
- Second, this thesis represents a preliminary demonstration of a comparatively low-cost resonance fatigue monitoring setup – a miniature electrodynamic shaker, a single scanning LDV, and a standard clamping fixture, assembled originally as supporting infrastructure for a planned quantum photonic vibrometer benchmark (Section 1.3 of Chapter 1) rather than as a dedicated, purpose-built fatigue rig. That this setup nonetheless produced a usable, physically consistent damage signal for at least one indicator (Section 6.1), even where none of the six DI formulations built on top of that signal reached a resolved population-level prognostic result (Section 6.4.1), and clean, classical fractographic evidence of fatigue as the governing mechanism (Section 6.6), suggests real potential for this class of low-cost setup, even though that potential was not fully realised here.
- Third, and more generally, the chapter's diagnostic throughline – tracing each limitation encountered (noisy modal identification, an unreliable CAE, a poor S-N fit, a narrow prognostic window) back to a specific, identifiable experimental or methodological cause rather than leaving it as unexplained scatter – is itself a transferable contribution: a future researcher repeating this kind of campaign now has a concrete list of what to prioritise (a demonstrably rigid boundary condition, a verified pure-bending load path, an adequately sized specimen population, N4SID over LSCF for

closely-spaced or highly-damped modes, and modal-identification-independent indicators over frequency and damping where data quality cannot be guaranteed in advance) that this thesis did not have available to it at the outset.

Read together, this thesis's contribution is best summarised not as a demonstration that modal-parameter-based fatigue monitoring works well under resonance excitation, but as a rigorous, traceable account of the conditions under which it does and does not, obtained from a genuinely integrated pipeline and checked at every stage against independent evidence. Chapter 7 states the conclusions for each research question in full and translates this diagnostic throughline into concrete practical recommendations.

Conclusions and Recommendations

This chapter closes the thesis by summarising the main findings, stating the conclusions reached with respect to each of the four research questions, identifying the study’s contribution and limitations, and outlining recommendations for both practical application and future research. Where Chapter 6 interpreted individual results in their immediate context, this chapter steps back to state, in each case, what is now known with reasonable confidence and what remains open.

7.1. Summary of Main Findings

This thesis set out to characterise how a specimen’s vibration response evolves under fixed-frequency resonance fatigue loading, to identify which indicators are most sensitive to early damage, to determine whether those trends can support reliable remaining-life prognosis, and to contextualise the fixed-frequency resonance protocol against the broader fatigue testing landscape. The campaign instrumented eight Al 6061-T6 beam specimens (Group C1) through complete fatigue life with a scanning LDV, supplemented by two specimens (Group C2) tested under a deliberately interrupted protocol with independent optical fractography validation, and by an exploratory autoencoder-based health indicator (Layer 0.5) and a dedicated S-N characterisation module developed after the core pipeline was established.

Across the eight Group C1 specimens, the measured vibration response proved considerably less reliable than the CDM framework’s idealised picture: only two of the three intended modes could be tracked with acceptable stability, and frequency, damping, coherence, and peak FRF amplitude – all dependent on clean modal identification – were substantially degraded by imperfect boundary-condition fixity and a load path that plausibly included an unintended axial component alongside the intended bending excitation. By contrast, HDI, amplitude RMS, and force RMS, none of which require modal parameter extraction, showed a consistent, broadly monotonic progression across most specimens. This same split reappeared, independently, in the Research Question R2 sensitivity ranking: HDI and mode-shape curvature dominated the SHAP-based importance ranking, while frequency and damping sat at the bottom, and the Spearman ranking – dominated instead by spatial features for reasons traced to what a purely monotonic-correlation measure rewards rather than to genuine damage sensitivity – disagreed with SHAP in informative rather than merely noisy ways. The Layer 0.5 autoencoder-based indicators, evaluated independently via leave-one-specimen-out cross-validation, achieved an AUROC at or below chance level (0.469 and 0.422) that remained essentially unchanged across three tested bottleneck sizes, pointing to a data-quality rather than an architectural explanation for the CAE’s poor generalisation. For Research Question R3, three of the six damage index formulations (D_{comp} , D_{PC1} , D_{CAE}) showed the cleanest within-specimen behaviour, yet none of the six formulations – D_{freq} included, despite its direct tie to the underlying CDM stiffness-loss relationship – reached a resolved population-level prognostic horizon within observed specimen life; confidence in any individual remaining-useful-life estimate remained low throughout, though longer-lived specimens consistently yielded more usable estimates, for reasons traced to both the additional data they provided and their milder rate of damage

evolution. For Research Question R4, the quantitative S-N characterisation was weak ($R^2 = 0.0227$, five distinct stress levels detected rather than the three the campaign was designed around), a result attributable to a first-order, linear-elastic force-to-stress calibration and the same plausible mixed loading state noted above, whereas the qualitative evidence that the fixed-frequency protocol sustains genuine resonant amplification throughout the test – rather than losing it as the natural frequency drifts – was considerably more solid. Finally, the independent optical fractography validation on the two Group C2 specimens found that the pipeline's damage alarms did coincide with real, physically observable surface change, but only once that change was already visually advanced rather than at its earliest, most subtle stage; classical fractographic evidence (beach marking, a shininess-versus-life-fraction relationship in fracture surface roughness, and a clearly demarcated final-fracture zone) independently confirmed fatigue as the dominant failure mechanism across the campaign as a whole.

7.2. Conclusions per Research Question

7.2.1. R1 -- Vibration Response Evolution

Research Question R1: How does the vibration response evolve as fatigue damage accumulates under resonant vibration loading?

Modal parameters evolved consistently with accumulating fatigue damage only where they did not depend on modal parameter extraction: HDI, amplitude RMS, and force RMS evolved in a manner consistent with progressive damage across most specimens, while frequency, damping, coherence, and peak FRF amplitude did not evolve reliably enough to support a confident CDM-consistency judgement, degraded by imperfect boundary-condition fixity and a plausible mixed bending-and-axial load path rather than by any property of the fatigue process itself. No confident conclusion can be drawn on whether a CDM-predicted hardening plateau occurred, since the frequency record's noise precludes distinguishing a genuine transient stiffening from measurement scatter. Spatial indicators did not unambiguously confirm notch-root localisation: MAC and mode-shape curvature broadly matched expectation, but curvature evaluated specifically at the notch scan point was erratic rather than consistently localised, so this thesis's spatial evidence should be read as consistent with, rather than as direct confirmation of, notch-initiated damage localisation.

7.2.2. R2 -- Indicator Sensitivity Ranking

Research Question R2: Which vibration-based indicators are most sensitive to early fatigue damage?

No single indicator is top-ranked across all three analytical methods, and this disagreement is itself the substantive conclusion rather than a weakness to be resolved: HDI and mode-shape curvature dominate the SHAP-based ranking, while Spearman correlation is instead dominated by spatial features for reasons traced to what a purely monotonic-correlation measure rewards (any consistent drift with life fraction, not necessarily damage-specific), not to a genuine disagreement about which features are most damage-sensitive. For a future monitoring system, HDI and curvature are the more defensible choice of primary indicator, since their SHAP-based importance is corroborated by the modal-identification-independence argument established for R1, whereas high Spearman rank alone is not sufficient evidence of damage sensitivity. The redundancy analysis (via PC1 loading) does not support a claim of extensive feature redundancy: variance is concentrated almost entirely in HDI and its first difference, with every other feature contributing negligibly to the dominant component, suggesting most of the 33-feature set is individually weak or noisy in this dataset rather than duplicating a small set of strong signals. The Layer 0.5 autoencoder-based indicators offered no meaningful advantage over the hand-engineered set: both HI1 and HI2 achieved leave-one-specimen-out AUROC at or below chance level (0.469 and 0.422), a result that held across three tested bottleneck sizes and is attributed to the small, noisy specimen population rather than to an architectural shortcoming.

7.2.3. R3 -- Remaining Useful Life Estimation

Research Question R3: Can vibration response trends be used to estimate remaining fatigue life?

The answer to R3 is qualified, and negative on the specific question of population-level prognosis within this campaign: vibration response trends can support remaining-life prognosis in principle, via

Method 1's single-specimen trend extrapolation, available from the first block of any new specimen without requiring a prior population. But none of the six damage index formulations tested reached a resolved population-level prognostic horizon (the $\Delta_{CI} < 0.30$ criterion applied to Method 2, the population Gaussian Process) within observed specimen life. This negative result is weaker than it first appears in one important respect: for three of the six formulations the uncertainty propagation itself diverged, so the criterion could not have been met at any block irrespective of the underlying damage signal (Section 6.4.1), and the result for those three is uninformative about detectability rather than evidence against it. For the formulations whose uncertainty estimates remained finite the conclusion stands as stated – D_{freq} included, despite its direct tie to the CDM stiffness-loss relationship and despite D_{comp} , D_{PC1} , and D_{CAE} showing cleaner within-specimen trajectories. The population Gaussian Process (Method 2) did not provide a uniformly demonstrated accuracy improvement over single-specimen trend extrapolation (Method 1) at this population size: it converged to a usable, resolved estimate for some specimen/formulation combinations but not others, consistent with a population of eight specimens providing enough shared structure to support the method only intermittently rather than reliably, and consistent with no formulation reaching the population horizon outright. This conclusion should be read as a proof of concept rather than a population-validated capability: with $n = 8$ specimens, none of the individual prognostic horizon, method-comparison, or formulation-ranking results should be generalised beyond this campaign without further population-scale validation.

7.2.4. R4 -- Resonance Testing in Context

Research Question R4: How does resonance-based vibration fatigue testing compare to traditional fatigue tests?

Resonance-based fatigue testing retains its expected qualitative advantage over conventional stress-controlled testing in accumulating fatigue cycles quickly by continuously exciting the specimen at or near a natural frequency, consistent with the established literature on accelerated resonance testing; this campaign's own block-cycle accounting is consistent with that expectation but does not itself provide a quantified efficiency comparison against a conventional fatigue test, since no such conventional-protocol comparison was run in parallel. The quantitative S-N characterisation obtained here is weak and should not be treated as a validated stress-life curve for this specimen geometry: the fitted Basquin relationship explains essentially none of the variance in fatigue life ($R^2 = 0.0227$), and the per-specimen force-to-stress calibration spread the eight specimens across five distinct stress levels rather than the three the campaign was designed around, a direct consequence of the calibration's linear-elastic simplification rather than of the underlying material behaviour. The lead-time overlay demonstrated that a resolved remaining-life estimate can in principle be obtained before fracture using this protocol, but only for a single specimen (C1-05, under two of the six DI formulations, both giving an identical 75,000-cycle lead time) rather than across the population, so this remains a proof-of-concept demonstration rather than evidence of population-wide advance warning; the qualitative case for the fixed-frequency protocol as a monitoring platform – that resonant amplification is sustained throughout the test rather than lost as the natural frequency drifts – is the more solid of this research question's two lines of evidence. A direct, controlled comparison between the fixed-frequency protocol used here and a tracked-frequency alternative was not performed in this thesis and remains an open question, addressed further in Section 7.5.

7.2.5. Other Contributions

Three further outcomes fall outside the four research questions.

- **An independent, vibration-free validation route for the detection logic.** The interrupted Group C2 protocol, in which fatigue loading is paused at the first live-monitoring alarm for optical fractography and then resumed to fracture, establishes that the alarm corresponded to real physical damage rather than to a measurement artefact. The method is transferable to any campaign that needs to separate the two, and requires no instrumentation beyond a microscope.
- **Quantified design requirements for resonance fatigue fixturing.** The pre-campaign characterisation batteries, the load configuration study, and the post-campaign diagnostic groups together yield a set of setup requirements that are stated as practical recommendations in Section 7.3: how to characterise a vibrometer installation before committing to a campaign, how to

verify a pure bending load path and a repeatable clamped boundary, and how to select force levels and block sizes so that specimens driven hardest still yield enough observations to analyse.

- **Three reusable software tools.** The LDV Analysis Suite, the Synthetic Data Toolkit implementing the coupled CDM–Basquin prediction, and the five-layer analysis pipeline with its live-monitoring path and S-N module, all built on a common HDF5 schema and released publicly with their documentation (Appendix D). These are usable independently of how this campaign’s results turned out, and the layer structure allows individual stages to be substituted or extended.

Taken with the answers to R1–R4 above, the contribution of this thesis is a traceable account of where vibration-based fatigue prognosis succeeds under controlled resonance conditions, where it does not, and which of its failures are attributable to the method rather than to the principle.

7.3. Practical Recommendations

The results support several concrete recommendations for the design of future modal-parameter-based fatigue monitoring systems, independent of whether such a system is deployed in a laboratory qualification setting or adapted toward in-service structural health monitoring.

7.3.1. Indicator and Feature Set Selection

For earliest possible detection and for indicator robustness under realistic, imperfect experimental conditions, HDI and amplitude RMS are the recommended primary indicators: both are computed directly from the raw response and excitation signals without requiring modal parameter extraction, and both showed a consistent, broadly monotonic progression across most specimens even where the modal-identification-dependent indicators (frequency, damping, coherence, peak FRF amplitude) did not (Section 6.1). For population-level prognosis specifically, none of the six formulations reached a resolved population-level prognostic horizon in this campaign (Section 6.4.1), so no formulation can currently be recommended on demonstrated performance; frequency shift remains the more theoretically motivated candidate for future population-level work, precisely because of its direct tie to the CDM stiffness-loss relationship rather than because of its per-specimen cleanliness (Section 6.3.1), but this is a recommendation for where to prioritise future validation effort, not a claim that it already works. These are not the same recommendation, and a future system should track a small, complementary set rather than a single indicator: HDI and amplitude RMS for robust, early anomaly flagging, and frequency shift as the priority candidate for population-level RUL estimation once validated, with the explicit caveat that frequency’s usefulness for the latter purpose depends on first resolving the boundary-condition and load-purity problems identified in Section 7.4.2, since a noisier frequency record than this campaign’s would be expected to degrade this candidate formulation further still. On redundancy specifically, this campaign’s data support dropping at most one clearly duplicated pair rather than a large fraction of the feature set: $\Delta\text{MAC}_{\text{model}}$ and $\text{MAC}_{\text{model}}$ were flagged as carrying near-identical information ($|r_{ij}| > 0.90$, Section 5.3), so a future instrumentation design could reasonably drop the first-difference variant and retain the raw MAC value, which requires one fewer processing step to compute; beyond this specific pair, most of the 33-feature set contributed negligibly to the dominant PC1 direction (Section 6.2.2), which speaks to those features’ individual weakness in this dataset rather than to a broader redundancy that would justify a large cut to the acquisition budget.

7.3.2. Damage Index Formulation and Prognosis Method

D_{freq} is the recommended formulation to prioritise for future prognostic validation, not because it was the best-behaved formulation on a per-specimen basis in this campaign – D_{comp} and D_{PC1} were both cleaner in that specific sense (Section 6.3.1) – and not because it was demonstrated to reach a resolved population-level prognostic horizon here, since none of the six formulations did. The recommendation rests instead on its direct tie to the CDM stiffness-loss relationship, which gives a principled, physically-grounded reason to expect it to generalise better across future campaigns than a formulation optimised only for within-specimen fit quality, once the boundary-condition and load-purity problems identified in Section 7.4.2 are resolved. The two autoencoder-derived formulations are not recommended for prognostic deployment at this population scale: both showed leave-one-specimen-out AUROC at or below chance level, a result attributed to insufficient, noisy training data rather than to an architectural limi-

tation (Section 6.2), and this should be revisited only once a substantially larger specimen population is available. On method choice, a hybrid deployment strategy is recommended over choosing one method exclusively: Method 1 (single-specimen trend extrapolation) requires no prior specimen population and should be used from the first block of any new campaign, while Method 2 (the population Gaussian Process) should be introduced once a prior population of at least five to six specimens is available (Section 4.4.3), since this is the practical minimum below which the GP cannot be expected to offer a reliable advantage over Method 1 alone – a real deployment prerequisite that a future system design should budget for explicitly rather than assume away.

7.3.3. Protocol Design: Fixed-Frequency vs. Tracked-Frequency

The fixed-frequency protocol is recommended when the primary objective is early damage detection and monitoring-oriented prognosis, and the tracked-frequency alternative is recommended when the primary objective is precise classical S-N characterisation. This follows directly from the trade-off underlying R4: a fixed drive frequency lets resonance detuning encode itself directly into phase and amplitude, amplifying the damage signal available to a monitoring system, but at the cost of allowing stress amplitude to vary uncontrolled over the test rather than holding it fixed as an S-N characterisation requires (Section 6.5.2). A test engineer should therefore treat this as a genuine either/or design decision made at the outset, rather than expecting one protocol to serve both purposes well simultaneously: this campaign's own S-N characterisation was weak in large part because the fixed-frequency protocol used was not designed to hold stress amplitude constant, and a future campaign whose primary goal is S-N curve generation should not adopt the fixed-frequency protocol on the expectation that a usable S-N curve will follow as a by-product.

7.3.4. Live Monitoring and Validation Practice

The interrupted-specimen validation track used in this campaign was informative and is recommended for continued use in future campaigns, but with two specific expansions motivated directly by this campaign's own limitations. First, a future campaign should interrupt each validation specimen at multiple checkpoints across its life, rather than only once at the first statistical alarm: this campaign's single-interruption design could confirm that the alarm coincided with real, visible damage, but could not establish whether the underlying indicators were also responding to damage before it became visible, since that requires a comparison at an earlier checkpoint that this protocol did not provide (Section 6.6.1). Second, the exclusion of the autoencoder-based indicators (HI1, HI2) from the live-monitoring alarm-triggering logic should remain in place for future deployments until a substantially larger training population is available, consistent with the near-chance validation performance found here and the data-hungriness this implies (Section 6.2); the two indicators can continue to be logged for post-hoc analysis without being trusted as alarm triggers in the interim. A further, low-cost addition is recommended based on an informal but genuinely useful observation from this campaign: an event camera recording the specimen during testing captured direct visual evidence of crack formation and propagation timing (Section 6.6.1) at negligible additional experimental cost, and a future campaign would benefit from incorporating this as a planned, continuous monitoring channel rather than as an opportunistic addition, since it provides exactly the kind of independent, high-temporal-resolution timing evidence that a single-checkpoint microscopy protocol cannot.

7.3.5. Designing a Vibration Fatigue Test Setup

The pre-campaign and load configuration process developed in this thesis (Sections 3.4.1–3.4.3) is itself a transferable design protocol for setting up a fixed-frequency resonance fatigue monitoring campaign on a new specimen geometry, and is summarised here as a practical recommendation independent of the specific aluminium beam results.

Synthesising this campaign's own design decisions, together with what its execution problems reveal in hindsight, into a short, ordered checklist for a future test engineer setting up a similar rig from scratch:

1. *Sensor characterisation first.* Establish the LDV noise floor and minimum detectable amplitude (T1–T2) before any specimen-specific work, since these set the detection sensitivity ceiling for everything that follows.

2. *Attachment point selection.* Choose the shaker attachment location so that it excites the target mode adequately while keeping the intended damage-initiation site (e.g. a notch) in a region of non-zero bending moment, rather than defaulting to the point of maximum displacement or maximum stress (Section 3.4.3).
3. *Verify a pure bending load path before the campaign begins, not after.* This campaign's own evidence – the fracture pattern of a mounting component consistent with mixed bending-and-axial loading rather than pure transverse motion (Section 6.1.1) – indicates that a nonlinear-geometry attachment can transmit an unintended axial force component at larger excitation amplitudes. A future setup should verify, at the highest planned excitation amplitude and before committing to the main campaign, that the force transmitted to the specimen remains predominantly transverse; this was not verified in advance here and is a direct, identified cause of this campaign's weak S-N characterisation (Section 6.5.1).
4. *Clamping and boundary condition verification.* Specify a fixed clamping torque, and, in light of Section 7.4.2, add an explicit boundary-condition re-verification step at the end of each fatigue test, not only before Phase 4 begins, so that genuine material scatter can be separated from fixture drift in future campaigns.
5. *Specimen manufacturing consistency as a controlled variable.* The Group D2 comparison (Section 6.1.7) found that, with the fixture held constant across four specimens, the largest source of variation in FRF repeatability was the specimens themselves: those produced on the intended equipment overlaid closely across runs, while those produced when it was unavailable showed coherence collapsing at the first bending mode and visible run-to-run scatter across the band. Specimen manufacturing should therefore be treated as a controlled variable of the campaign rather than a preliminary to it. In practice this means machining the full population on the same equipment in a single batch where possible, recording any deviation from that as a specimen attribute, and screening each specimen with a short repeatability check before it enters the campaign, so that a specimen whose baseline FRF cannot be reproduced across successive scans is identified before fatigue loading rather than after.
6. *Force level selection via a pilot study.* Apply the three-criterion approach used for Group B (Section 3.4.3) – fracture location consistency, an acceptable block-count range for adequate temporal resolution, and a minimum force coherence threshold across the modes of interest – to a small pilot population before committing to the main campaign's force levels.
7. *Design the stress-level matrix up front, and calibrate force-to-stress before, not after, testing.* Decide the number of stress levels and repeats per level (Section 7.4.1) at the design stage, trading off population size against distinct-level count explicitly, since the monitoring objective (R1–R3, favouring more specimens per condition) and the S-N objective (R4, favouring more distinct levels) draw on the same finite specimen budget; and complete any force-to-stress finite element calibration (Section 6.5.1) before the campaign, using a model whose modal characteristics have been checked against a representative specimen, rather than treating it as a follow-up step once testing is already complete.

7.4. Limitations of the Study

The following limitations qualify the conclusions stated above and should be read alongside them. They are grouped into four themes rather than by chapter, since several individual limitations span multiple research questions.

7.4.1. Experimental Campaign Scale

The Group C1 population of eight specimens is small relative to what population-level prognostic validation ordinarily requires, and the Group C2 validation pair smaller still. Every population-level result in this thesis carries that constraint: the leave-one-out cross-validation trains on seven specimens, the Gaussian Process has seven to generalise from, and the S-N fit has eight points across five effectively distinct stress levels, since the achieved drive amplitudes varied enough within each intended level to separate under the fit's own gap rule.

The protocol compounds this in a way the campaign design did not anticipate. Because fatigue blocks

are of fixed size, the number of observations a specimen yields is set by how long it survives, so the specimens driven hardest yield the fewest. S01–S03 completed one or two Phase 4 blocks each, too few to fit the two-parameter logistic underlying Layer 2. Every damage-index and remaining-life result therefore rests on five specimens, and the uncertainty analysis on four, rather than the nominal eight. Temporal resolution and stress level are coupled under this protocol, and Section 7.5.1 proposes a block-sizing scheme that decouples them.

7.4.2. Measurement Reliability and Calibration

Section 6.1 of Chapter 6 establishes that measurement reliability, rather than indicator choice, governed what this campaign could resolve. Four limitations follow from it.

First, only two of the three intended modes could be tracked with acceptable stability, and the MAC diagnostic shows that mode-shape repeatability between two undamaged blocks of the same specimen is poor enough to leave the spatial indicators little dynamic range, and that the mode labels do not reliably track the same physical mode between blocks. Every per-mode trajectory in this thesis is therefore a trajectory of whatever the identification routine assigned that label in each block. This is the primary reason the indicators requiring no modal identification are recommended over those that do (Section 7.3.1).

Second, the clamping arrangement and the shaker attachment introduced amplitude-dependent effects that the protocol did not control, evidenced directly by physical inspection of both. The load cell records only the component along the intended axis, so any parasitic moment is applied to the specimen but never measured.

Third, the force-to-stress calibration underlying the R4 characterisation is first-order rather than validated. The finite element analysis assumed linear-elastic behaviour for a geometry whose response is not linear at the stress concentrations involved, the model's own modal characteristics differed substantially from the physical specimens' in some cases, and although a separate analysis was run for every specimen it produced only three distinct values across the eight, the geometric differences between six of them being too small for the model to resolve.

Fourth, the nine-point scan grid falls below the density recommended for reliable modal curvature estimation by finite differences, and does not span the notch, so the spatial indicators detect damage through its effect on the deflected shape nearby rather than by imaging the crack.

7.4.3. Methodological Scope

Two aspects of the analytical methodology were deliberately scoped narrowly and remain open. First, this thesis does not perform a literal side-by-side comparison between the fixed-frequency protocol used throughout the campaign and a phase-locked, tracked-frequency protocol on an equivalent specimen population (Section 2.10.3 of Chapter 2); the amplification argument for fixed-frequency monitoring sensitivity is supported by direct evidence from this campaign (Section 6.5.2) but not tested against a tracked-frequency control group. Second, the modal feature extraction used throughout this thesis is built on linear modal analysis: the identification methods, the FRF-based damage indicators, and the CDM stiffness-loss relationship all assume linear or quasi-linear structural response at each measurement instant, with HDI the only indicator that explicitly captures a nonlinear effect (crack-breathing super-harmonics), and only as a scalar summary rather than through a full nonlinear modal characterisation.

7.4.4. External Constraints

Two further limitations arise from constraints external to the analytical methodology itself. The optical fractography validation (Section 5.7) was deliberately scoped as qualitative, providing discrete, vibration-free checkpoints rather than a continuous, quantitative crack-length record; sub-surface or microscopic damage below the resolution of the optical microscope used would not be detected, so the "already visually advanced" character of the observed damage (Section 6.6.1) reflects the resolution and single-checkpoint timing of this protocol rather than a demonstrated absence of earlier, sub-visible sensitivity. Separately, as described in Chapter 1, this thesis's research direction was reoriented after the quantum photonic vibrometer (QPV) instrument that was the original basis for the project became

unavailable due to equipment failure; the resulting fatigue monitoring study is well-matched to the available experimental infrastructure (Section 2.10.4), but the originally intended comparison between LDV-based and QPV-based vibration measurement for fatigue testing applications was not performed and remains a separate, unaddressed research question.

7.5. Suggestions for Future Research

7.5.1. Improving Measurement Reliability and Calibration

Three concrete improvements follow directly from the measurement and calibration limitations identified in Section 7.4.2. These are placed first because they are prerequisites: adding specimens to a protocol whose boundary condition, load path, and stress calibration remain uncontrolled would multiply the existing scatter rather than resolve it. First, resolving the boundary-condition and load-purity problems that limited modal identification to two of three intended modes – specifically, the rigid clamping and verified pure-bending load path recommended in Section 7.3.5 – should be treated as a prerequisite for any repeat campaign, since a third reliably-tracked mode would directly expand the spatial and point-tier feature set available to Layers 1–3. Second, an end-of-test boundary-condition re-verification step, rather than only a pre-Phase 4 check, would let a future campaign separate genuine material and surface-condition scatter from fixture drift in the population-level statistics, which the present campaign could not do retrospectively. Third, and most directly relevant to completing the R4 contribution, the force-to-stress calibration should be redone with a finite element model whose own modal characteristics have been validated against a representative physical specimen before it is trusted to produce per-specimen k values (Section 7.4.2), and using a geometrically nonlinear analysis rather than a linear-elastic one given the notch stress concentration involved. Only once this calibration is on a validated footing would the additional repeats per stress level suggested in Section 7.5.2 translate into a meaningful, quantitative comparison against the published Al 6061-T6 S-N literature.

7.5.2. Expanding the Experimental Campaign

Once the measurement and calibration improvements above are in place, the most direct extension of this work is a larger specimen population, addressing the small- n limitation (Section 7.4.1) directly. A population of $n \geq 20$ specimens, tested under the same fixed-frequency protocol, would substantially tighten the confidence intervals on the R2 sensitivity ranking, provide a more robust Gaussian Process training set for R3 (allowing, for example, a held-out test set distinct from the leave-one-out training pool), and allow the redundancy and PC1 structure to be validated rather than estimated from a single eight-specimen sample. Expanding the Group C2 validation track to more than two specimens would similarly strengthen the fractography validation from indicative to statistically supported. A larger population alone would not, however, resolve the coupling between drive level and temporal resolution identified in Section 7.4.1: because blocks are of fixed size, the specimens driven hardest still yield the fewest observations. Scaling the block size inversely with the expected life at each stress level, so that every specimen contributes a comparable number of blocks regardless of how quickly it fails, would give the Layer 2 and Layer 3 fits a usable record from the short-lived specimens that the present protocol excludes. On the R4 side specifically, increasing from two to three repeats per stress level to five or more, and adding one or two further stress levels at the extremes of the range validated by the Group B study, would tighten the Basquin slope confidence interval (Section 7.4.1) substantially without requiring any change to the experimental protocol itself, though, as Section 7.5.1 below notes, this is only worth pursuing once the calibration limitation underlying the current fit is addressed first.

7.5.3. Methodological Scope: Protocol Comparison, CAE Integration, and Nonlinear Modal Analysis

Two extensions would directly close the methodological scope limitations identified in Section 7.4.3. First, running the fixed-frequency protocol against a phase-locked, tracked-frequency control group on an equivalent specimen population, with the same indicator sensitivity and prognostic horizon analysis applied to both, would resolve whether the apparent sensitivity advantage reported in Section 6.5.2 is a genuine effect of the protocol or an artefact of this campaign's specimens and loading history. Second, the linear-response assumption noted there could be relaxed by drawing on nonlinear modal analysis techniques: backbone curve tracking, which characterises the amplitude-dependent natural

frequency and damping relationship and is directly relevant to the phase-locked-loop methods above, or a fuller characterisation of the crack-breathing bilinear stiffness behaviour. Both could in principle extract more information from the raw FRF data already being collected, without requiring any change to the experimental protocol.

The Layer 3 Gaussian Process model has since been extended to accept D_{CAE} and $D_{\text{CAE,latent}}$ as inputs alongside the four hand-engineered formulations, resolving what was an open methodological gap at the time this campaign's analysis began; the resulting like-for-like Method 1/Method 2 comparison for all six formulations (Chapter 5) confirms that the autoencoder-derived formulations remain no more reliable than the hand-engineered ones at this population scale, consistent with the recommendation in Section 7.3.2 against deploying them until a substantially larger training population is available (Section 7.5.2).

Beyond these three, a larger and cleaner dataset arising from the improvements above would also open several broader machine-learning extensions worth noting, though at a lower priority than resolving the scope limitations directly: alternative autoencoder architectures (attention-based or variational formulations benchmarked against the current convolutional baseline); a sequence-aware model for the DI trajectory, since the current Layer 3 treatment is essentially memoryless at each prediction checkpoint and a recurrent or transformer-based model trained on the full per-specimen history could exploit trajectory shape rather than only the instantaneous value; transfer learning across specimen geometries, testing whether a model trained on one geometry or material transfers with limited fine-tuning; and a joint or multi-task formulation spanning detection, ranking, indexing, and prognosis rather than treating them as four separate stages. Each of these builds directly on the pipeline architecture already established (Section 4.1 of Chapter 3) rather than requiring a new framework.

7.5.4. External Constraints: Fractography Depth and the QPV Comparison

Two extensions follow from the external constraints identified in Section 7.4.4. First, beyond simply expanding the Group C2 population (Section 7.5.2), the fractography protocol itself would benefit from interrupting each validation specimen at multiple checkpoints across its life rather than only once at the first statistical alarm (Section 7.3.4), which is the only way to test whether the pipeline's indicators respond to damage before it becomes visible at the resolution used here, a question this campaign's single-checkpoint design could not resolve. Second, as noted in Chapter 1, the original motivation for this project was a comparison between the scanning LDV and a quantum photonic vibrometer (QPV) for fatigue testing applications. With the fatigue monitoring pipeline, specimen protocol, and analysis framework now established by this thesis, repeating a subset of the fatigue campaign with a functioning QPV instrument, once available, would allow a direct comparison of measurement noise floor, spatial resolution, and damage indicator sensitivity between the two measurement technologies on the same specimen geometry and loading protocol, completing the original research direction that this thesis's topic change was forced to set aside.

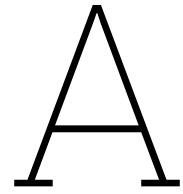
References

- Allemang, R. J. (2003). The modal assurance criterion — twenty years of use and abuse. *Sound and Vibration*, 37(8), 14–20.
- Allemang, R. J., & Avitabile, P. (2022). *Handbook of experimental structural dynamics*. Springer. <https://doi.org/10.1007/978-1-4939-6503-8>
- Allemang, R. J., & Brown, D. L. (1982). A correlation coefficient for modal vector analysis. *Proceedings of the 1st International Modal Analysis Conference (IMAC)*, 110–116.
- Andreaus, U., & Baragatti, P. (2012). Experimental damage detection of cracked beams by using non-linear characteristics of forced response. *Mechanical Systems and Signal Processing*, 31, 382–404. <https://doi.org/10.1016/j.ymssp.2012.04.007>
- Andreaus, U., & Baragatti, P. (2009). Fatigue crack growth, free vibrations, and breathing crack detection of aluminium alloy and steel beams. *Journal of Strain Analysis for Engineering Design*, 44(7), 595–608. <https://doi.org/10.1243/03093247JSA527>
- ASTM International. (2021). Standard practice for conducting force controlled constant amplitude axial fatigue tests of metallic materials [ASTM E466-21]. *ASTM International*.
- Atamuradov, V., Medjaher, K., Camci, F., Zerhouni, N., Dersin, P., & Lamoureux, B. (2020). Machine health indicator construction framework for failure diagnostics and prognostics. *Journal of Signal Processing Systems*, 92, 591–609. <https://doi.org/10.1007/s11265-019-01491-4>
- Bandara, R. P., Chan, T. H. T., & Thambiratnam, D. P. (2014). Frequency response function based damage identification using principal component analysis and pattern recognition technique. *Engineering Structures*, 66, 116–128. <https://doi.org/10.1016/j.engstruct.2014.01.044>
- Bendat, J. S., & Piersol, A. G. (1993). *Engineering applications of correlation and spectral analysis* (2nd ed). Wiley.
- Carminati, M., & Ricci, S. (2018). Structural damage detection using nonlinear vibrations. *International Journal of Aerospace Engineering*, 2018, 1901362. <https://doi.org/10.1155/2018/1901362>
- Česnik, M., & Slavič, J. (2014). Vibrational Fatigue and Structural Dynamics for Harmonic and Random Loads. *Strojniški vestnik – Journal of Mechanical Engineering*, 60(5), 339–348. <https://doi.org/10.5545/sv-jme.2013.1831>
- Česnik, M., Slavič, J., & Boltežar, M. (2012). Uninterrupted and accelerated vibrational fatigue testing with simultaneous monitoring of the natural frequency and damping. *Journal of Sound and Vibration*, 331(24), 5370–5382. <https://doi.org/10.1016/j.jsv.2012.06.022>
- Coble, J., & Hines, J. W. (2009). Identifying optimal prognostic parameters from data: A genetic algorithms approach. *Annual Conference of the Prognostics and Health Management Society*.
- Conover, W. J. (1999). *Practical nonparametric statistics* (3rd). Wiley.
- Di Maio, D., & Magi, F. (2015). Development of testing methods for endurance trials of composites components. *Journal of Composite Materials*, 49(24), 2977–2991.
- Di Maio, D., Magi, F., & Sever, I. A. (2018). Damage Monitoring of Composite Components under Vibration Fatigue using Scanning Laser Doppler Vibrometer. *Experimental Mechanics*, 58(3), 499–514. <https://doi.org/10.1007/s11340-017-0367-y>
- Dirlik, T., & Benasciutti, D. (2021). Dirlik and Tovo-Benasciutti Spectral Methods in Vibration Fatigue: A Review with a Historical Perspective. *Metals*, 11(9), 1333. <https://doi.org/10.3390/met11091333>
- Eltouny, K., Gomaa, M., & Liang, X. (2023). Unsupervised Learning Methods for Data-Driven Vibration-Based Structural Health Monitoring: A Review. *Sensors*, 23(6), 3290. <https://doi.org/10.3390/s23063290>
- Ewins, D. J. (2000). *Modal testing: Theory, practice, and application* (2nd ed). Research Studies Press.
- Furman, B. A., German, J. E., Wagner, J. M., Heninger, J. B., & Berke, R. B. (2025). Accelerated fatigue measurements using a vibration assembly which loads two specimens simultaneously. *Journal of Applied Mechanics*, 92(10), 101001. <https://doi.org/10.1115/1.4068699>

- Gao, J., Wang, C., Xu, Z., Wang, J., Yan, S., & Wang, Z. (2022). Gaussian process regression based remaining fatigue life prediction for metallic materials under two-step loading. *International Journal of Fatigue*, 158, 106730. <https://doi.org/10.1016/j.ijfatigue.2022.106730>
- García-Fernández, N., Aenlle, M., Álvarez-Vázquez, A., Muniz-Calvente, M., & Fernández, P. (2023). A review on fatigue monitoring of structures. *International Journal of Structural Integrity*, 14(2), 133–165. <https://doi.org/10.1108/IJSI-09-2022-0117>
- George, T. J., Seidt, J., Shen, M.-H. H., Nicholas, T., & Cross, C. J. (2004). Development of a novel vibration-based fatigue testing methodology. *International Journal of Fatigue*, 26(5), 477–486. <https://doi.org/10.1016/j.ijfatigue.2003.10.012>
- Ghazimoghadam, S., & Hosseinzadeh, S. A. A. (2024). A novel unsupervised deep learning approach for vibration-based damage diagnosis using a multi-head self-attention lstm autoencoder. *Measurement*, 229, 114410. <https://doi.org/10.1016/j.measurement.2024.114410>
- Guan, X., Liu, Y., Saxena, A., Celaya, J., & Goebel, K. (2009). Entropy-based probabilistic fatigue damage prognosis and algorithmic performance comparison. *Annual Conference of the Prognostics and Health Management Society (PHM09)*.
- Harris, C. M., & Piersol, A. G. (Eds.). (2002). *Harris' shock and vibration handbook* (5. ed). McGraw-Hill.
- Hippold, P., Scheel, M., Renon, L., & Krack, M. (2024). Robust and fast backbone tracking via phase-locked loops. *Mechanical Systems and Signal Processing*. <https://doi.org/10.1016/j.ymssp.2024.111670>
- Huang, Y., Sua, Y., Rehai, P., Zhu, S., Tafone, D., & RAMANATHAN, J. (2023, September). *Approaches, apparatuses, and methods for non-interferometric quantum photonics vibrometry* (US20230288248A1). <https://patents.google.com/patent/US20230288248A1/en?q=Patent+US20230288248A1>
- International Organization for Standardization. (2017). Metallic materials — fatigue testing — axial force-controlled method [ISO 1099:2017]. ISO.
- Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, 374(2065), 20150202. <https://doi.org/10.1098/rsta.2015.0202>
- Jones, K., Musinski, W. D., Pilchak, A. L., John, R., Shade, P. A., Rollett, A. D., & Holm, E. A. (2023). Predicting fatigue crack growth metrics from fractographs: Towards fractography by computer vision. *International Journal of Fatigue*, 177, 107915. <https://doi.org/10.1016/j.ijfatigue.2023.107915>
- Karolczuk, A., & Słowski, M. (2022). Application of the gaussian process for fatigue life prediction under multiaxial loading. *Mechanical Systems and Signal Processing*, 167, 108599. <https://doi.org/10.1016/j.ymssp.2021.108599>
- Kuttner, T., & Rohnen, A. (2023). *Practice of Vibration Measurement: Measurement Technology and Vibration Analysis with MATLAB®*. Springer Fachmedien. <https://doi.org/10.1007/978-3-658-38463-0>
- Lasen Andrade, M., Di Maio, D., De Bono, D., & Peluzzo, M. (2025). An experimental investigation of fatigue damage growth in composites materials using the vibration response phase decay. *Experimental Techniques*. <https://doi.org/10.1007/s40799-025-00826-0>
- Lee, Y.-L. (2005). *Fatigue testing and analysis: Theory and practice*. Elsevier Butterworth-Heinemann. <https://doi.org/10.1016/B978-0-7506-7719-6.X5000-3>
- Lemaitre, J., & Chaboche, J.-L. (1990). *Mechanics of solid materials*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139167970>
- Li, L., Morgantini, M., & Betti, R. (2023). Structural damage assessment through a new generalized autoencoder with features in the quefrequency domain. *Mechanical Systems and Signal Processing*, 184, 109713. <https://doi.org/10.1016/j.ymssp.2022.109713>
- Liang, Z., Ramakrishnan, K. R., Ng, C.-T., Zhang, Z., & Fu, J. (2024). Vibration-based prediction of residual fatigue life for composite laminates through frequency measurements. *Composite Structures*, 329, 117771. <https://doi.org/10.1016/j.compstruct.2023.117771>
- Liu, F. T., Ting, K. M., & Zhou, Z.-H. (2008). Isolation forest. *2008 Eighth IEEE International Conference on Data Mining (ICDM)*, 413–422. <https://doi.org/10.1109/ICDM.2008.17>
- Lorenzino, P., & Navarro, A. (2015). The variation of resonance frequency in fatigue tests as a tool for *in-situ* identification of crack initiation and propagation, and for the determination of cracked areas. *International Journal of Fatigue*, 70, 374–382. <https://doi.org/10.1016/j.ijfatigue.2014.08.002>

- Lu, C. J., & Meeker, W. O. (1993). Using degradation measures to estimate a time-to-failure distribution. *Technometrics*, 35(2), 161–174. <https://doi.org/10.1080/00401706.1993.10485038>
- Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems*, 30. <https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html>
- Magi, F., Di Maio, D., & Sever, I. (2016a). Damage initiation and structural degradation through resonance vibration: Application to composite laminates in fatigue. *Composites Science and Technology*, 132, 47–56. <https://doi.org/10.1016/j.compscitech.2016.06.013>
- Magi, F., Di Maio, D., & Sever, I. (2016b). Scanning LDV Measurement Technology for Vibration Fatigue Testing. In J. De Clerck & D. S. Epp (Eds.), *Rotating Machinery, Hybrid Test Methods, Vibro-Acoustics & Laser Vibrometry, Volume 8* (pp. 87–94). Springer International Publishing. https://doi.org/10.1007/978-3-319-30084-9_8
- MatWeb, LLC. (2026). Aluminum 6061-t6; 6061-t651. <https://www.matweb.com/search/datasheet.aspx?MatGUID=b8d536e0b9b54bd7b69e4124d8f1d20a>
- Mei, H., Migot, A., Haider, M. F., Joseph, R., Bhuiyan, M. Y., & Giurgiutiu, V. (2019). Vibration-Based In-Situ Detection and Quantification of Delamination in Composite Plates. *Sensors*, 19(7), 1734. <https://doi.org/10.3390/s19071734>
- Mishra, S., Vanli, O. A., & Park, C. (2015). A multivariate cumulative sum method for continuous damage monitoring with lamb-wave sensors. *International Journal of Prognostics and Health Management*.
- Montalvão, D., Maia, N. M. M., & Ribeiro, A. M. R. (2006). A review of vibration-based structural health monitoring with special emphasis on composite materials. *Shock and Vibration Digest*, 38(4), 295–324. <https://doi.org/10.1177/0583102406065898>
- Murakami, S. (2012). *Continuum damage mechanics: A continuum mechanics approach to the analysis of damage and fracture*. Springer. <https://doi.org/10.1007/978-94-007-2666-6>
- Okur, F. Y., Altunışık, A. C., & Kalkan Okur, E. (2025). A novel approach for anomaly detection in vibration-based structural health monitoring using autoencoders in deep learning. *Structural Control and Health Monitoring*, 2025, 1–13. <https://doi.org/10.1155/stc/5602604>
- Page, E. S. (1954). Continuous inspection schemes. *Biometrika*, 41(1-2), 100–115. <https://doi.org/10.1093/biomet/41.1-2.100>
- PCB Piezotronics, Inc. (2026a). Model 288d01: Mechanical impedance sensor. https://www.pcb.com/contentstore/docs/pcb_corporate/vibration/products/manuals/288d01.pdf
- PCB Piezotronics, Inc. (2026b). Model k2007e01: Smartshaker with integrated amplifier. <https://www.pcb.com/products?m=K2007E01>
- Polytec GmbH. (2026). Psv-500 scanning vibrometer. <https://www.polytec.com/eu/vibrometry/products/full-field-vibrometers/psv-500-scanning-vibrometer>
- Proso, U., Slavič, J., & Boltežar, M. (2016). Vibration-fatigue damage accumulation for structural dynamics with non-linearities. *International Journal of Mechanical Sciences*, 106, 72–77. <https://doi.org/10.1016/j.ijmecsci.2015.12.005>
- Ramanathan, J., Garikapati, M., Jalata, I., Kadrollimath, C., Mahamuni, P., Hoffman, D., Satyanand, P., Vrahoretis, N., Sua, Y. M., & Huang, Y. (2025). Material recognition using single photon vibrometer. *Optics Express*, 33(13), 27003–27013. <https://doi.org/10.1364/OE.566439>
- Rao, S. S. (2017). *Mechanical vibrations* (6th). Pearson.
- Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian processes for machine learning*. MIT Press. <https://doi.org/10.7551/mitpress/3206.001.0001>
- Rehain, P., Ramanathan, J., Sua, Y. M., Zhu, S., Tafone, D., & Huang, Y.-P. (2021). Single-photon vibrometry. *Optics Letters*, 46(17), 4346–4349. <https://doi.org/10.1364/OL.433423>
- Riley, W. J. (2008). *Handbook of frequency stability analysis* (tech. rep. No. NIST Special Publication 1065). NIST.
- Rothberg, S., Allen, M., Castellini, P., Di Maio, D., Dirckx, J., Ewins, D., Halkon, B., Muyschondt, P., Paone, N., Ryan, T., Steger, H., Tomasini, E., Vanlanduit, S., & Vignola, J. (2017). An international review of laser Doppler vibrometry: Making light work of vibration measurement. *Optics and Lasers in Engineering*, 99, 11–22. <https://doi.org/10.1016/j.optlaseng.2016.10.023>
- Rytter, A. (1993). *Vibrational based inspection of civil engineering structures* [Doctoral dissertation, Aalborg University].

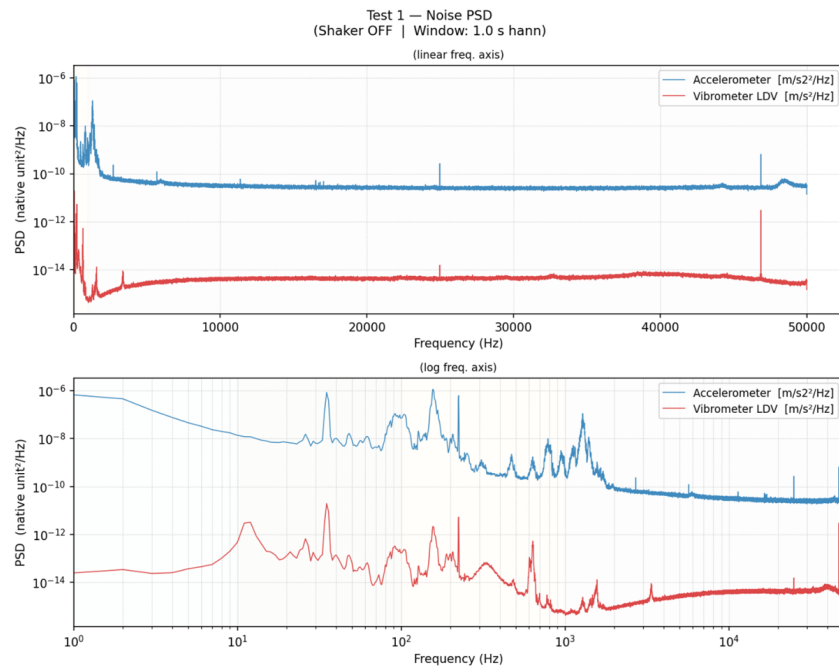
- Saxena, A., Celaya, J., Saha, B., Saha, S., & Goebel, K. (2008). Metrics for evaluating performance of prognostic techniques. *2008 International Conference on Prognostics and Health Management (PHM)*, 1–17. <https://doi.org/10.1109/PHM.2008.4711436>
- Saxena, A., Celaya, J., Saha, B., Saha, S., & Goebel, K. (2010). Metrics for offline evaluation of prognostic performance. *International Journal of Prognostics and Health Management*, 1(1). <https://doi.org/10.36001/ijphm.2010.v1i1.1336>
- Schijve, J. (2008). *Fatigue of structures and materials* (2nd ed). Springer. <https://doi.org/10.1007/978-1-4020-6808-9>
- Silitonga, S., Maljaars, J., Soetens, F., & Snijder, H. H. (2013). *Survey on damage mechanics models for fatigue life prediction*.
- Singh, T., & Sehgal, S. (2022). Damage identification using vibration monitoring techniques. *Materials Today: Proceedings*, 69, 133–141. <https://doi.org/10.1016/j.matpr.2022.08.204>
- Slavič, J. (2021). *Vibration fatigue by spectral methods: From structural dynamics to fatigue damage – theory and experiments*. Elsevier. <https://doi.org/10.1016/C2019-0-04580-3>
- Teng, Y., Xie, L., & Zhang, H. (2022). Experimental study on vibration fatigue behavior of aircraft aluminum alloy 7050. *Materials*, 15(21), 7555. <https://doi.org/10.3390/ma15217555>
- Tomasini, E. P., & Castellini, P. (2020). *Laser doppler vibrometry: A multimedia guide to its features and usage*. Springer. <https://doi.org/10.1007/978-3-662-61318-4>
- Toufigh, V., & Ranjbar, I. (2024). Unsupervised deep learning framework for ultrasonic-based distributed damage detection in concrete: Integration of a deep auto-encoder and isolation forest for anomaly detection. *Structural Health Monitoring*, 23(3), 1313–1333. <https://doi.org/10.1177/14759217231183143>
- Van Overschee, P., & De Moor, B. (1994). N4sid: Subspace algorithms for the identification of combined deterministic-stochastic systems. *Automatica*, 30(1), 75–93. [https://doi.org/10.1016/0005-1098\(94\)90230-5](https://doi.org/10.1016/0005-1098(94)90230-5)
- Vibrometer - quantum computing inc. (2018). *Quantumcomputinginc.com*. <https://quantumcomputinginc.com/products/r-and-d-offerings/vibrometer>
- Voudouris, G., Di Maio, D., & Sever, I. A. (2020). Experimental fatigue behaviour of CFRP composites under vibration and thermal loading. *International Journal of Fatigue*, 140. <https://doi.org/10.1016/j.ijfatigue.2020.105791>
- Wang, R.-J., Shang, D.-G., Li, L.-S., & Li, C.-S. (2008). Fatigue damage model based on the natural frequency changes for spot-welded joints. *International Journal of Fatigue*, 30(6), 1047–1055. <https://doi.org/10.1016/j.ijfatigue.2007.08.007>
- Welch, P. D. (1967). The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, 15(2), 70–73. <https://doi.org/10.1109/TAU.1967.1161901>
- Worden, K., Tomlinson, G., & Yagasaki, K. (2002). Nonlinearity in Structural Dynamics: Detection, Identification and Modeling. *Applied Mechanics Reviews*, 55(2), B26. <https://doi.org/10.1115/1.1451161>
- Worden, K., Farrar, C. R., Haywood, J., & Todd, M. (2008). A review of nonlinear dynamics applications to structural health monitoring. *Structural Control and Health Monitoring*, 15(4), 540–567. <https://doi.org/10.1002/stc.215>
- Xu, Y. M., Ding, K. Q., Cao, M. S., & Manoach, E. (2018). Bispectral dynamics features for characterizing structural fatigue damage. *Journal of Vibroengineering*, 20(5), 2073–2084. <https://doi.org/10.21595/jve.2018.19963>
- Yahr, G. T. (1997). Fatigue design curves for 6061-t6 aluminum. *Journal of Pressure Vessel Technology*, 119(2), 211–215. <https://doi.org/10.1115/1.2842286>



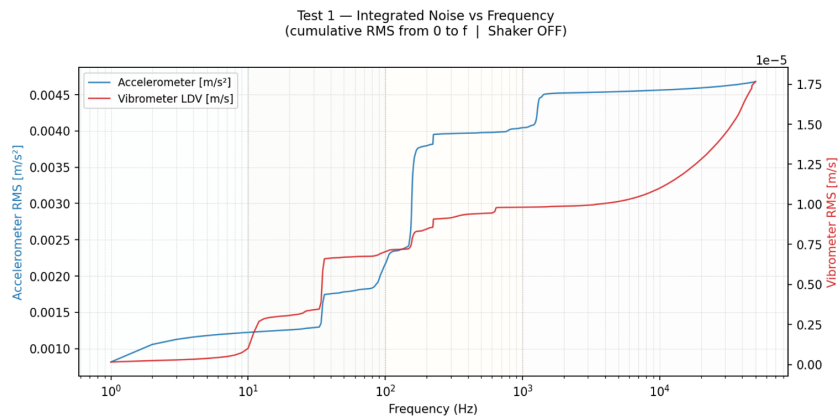
Sensor Characterisation Reports (T1--T4)

This appendix reproduces the full result set of the T1–T4 sensor characterisation battery described in Section 3.4.1. The tests characterise the intrinsic performance of the LDV and its drive chain, independently of any specimen: with the sole exception of T4d, no beam was mounted, and the LDV was pointed directly at the load cell and shaker armature. The consolidated outputs that propagate into the fatigue protocol are summarised in the main text; the per-test detail is given here.

A.1. T1 -- Noise Characterisation



(a) Noise power spectral density.



(b) Integrated (cumulative RMS) noise.

Figure A.1: T1 noise characterisation, shaker off, for both the LDV and the reference accelerometer on the load cell. (a) Noise PSD: the LDV noise floor sits several orders of magnitude below the accelerometer's across the full bandwidth. (b) Cumulative RMS noise from 0 to f : the LDV reaches approximately 1.0×10^{-5} m/s within the 0–1000 Hz band spanning the first three bending modes, rising to approximately 1.8×10^{-5} m/s across the full 50 kHz measurement bandwidth, consistently below the accelerometer's cumulative noise at every frequency.

A.2. T2 -- Minimum Detectable Amplitude

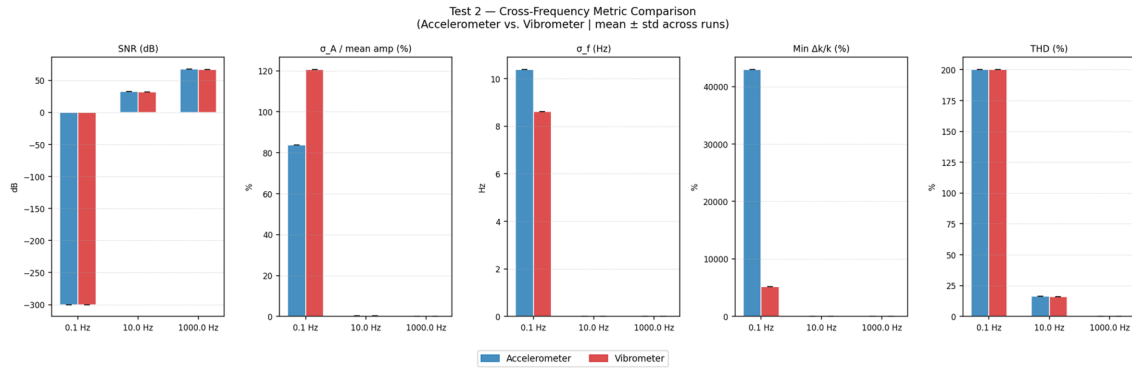


Figure A.2: T2 cross-frequency metric comparison (accelerometer vs. LDV): SNR, amplitude uncertainty, frequency uncertainty, stiffness-change resolution, and THD at 0.1, 10, and 1000 Hz, at the lowest achievable shaker drive voltage. The extreme values at 0.1 Hz (SNR near -300 dB, THD near 200%) reflect that the shaker was driven at a fixed *voltage* rather than a fixed force or displacement (Section 3.4.3): at 0.1 Hz the resulting displacement is far too small to produce a usable response at all, whereas by 10–1000 Hz the same drive voltage yields an adequately excited, well-behaved response for both sensors.

A.3. T3 -- Broadband and Harmonic Excitation Quality

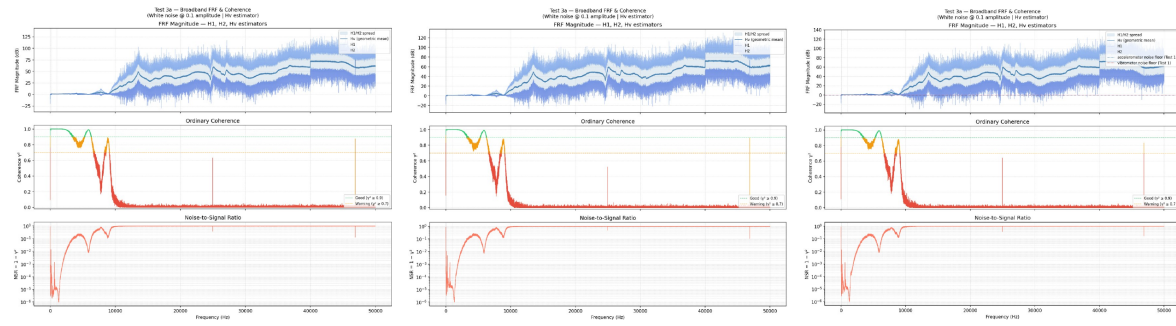


Figure A.3: T3a broadband excitation quality, repeated over three acquisitions under white-noise excitation at 0.1 amplitude.

Each column shows, top to bottom: FRF magnitude with the H_1/H_2 spread band and the H_v geometric mean, ordinary coherence γ^2 against the 0.9 “good” and 0.7 “warning” thresholds, and the corresponding noise-to-signal ratio. Coherence is close to unity below approximately 5 kHz and falls away above roughly 10 kHz, which bounds the usable band for FRF estimation well above the 0–500 Hz range within which the campaign’s tracked modes lie.

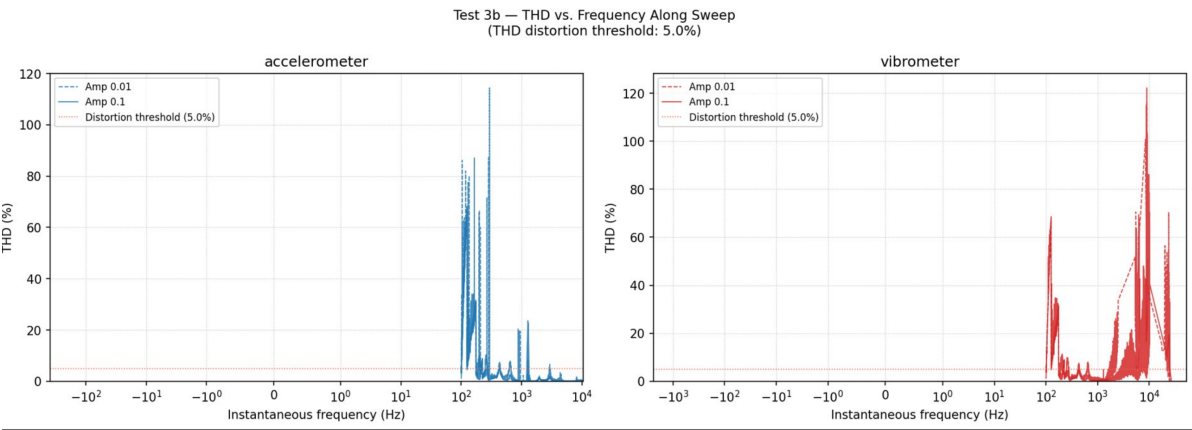


Figure A.4: T3b total harmonic distortion against instantaneous frequency along a swept-sine excitation, at two drive amplitudes, for the reference accelerometer (left) and the LDV (right). The 5% distortion threshold is marked. Both sensors remain below the threshold across the low-frequency region containing the campaign’s drive frequency, with distortion rising sharply only above approximately 100 Hz for the accelerometer and above 1 kHz for the LDV.

A.4. T4 -- Disturbance Sensitivity

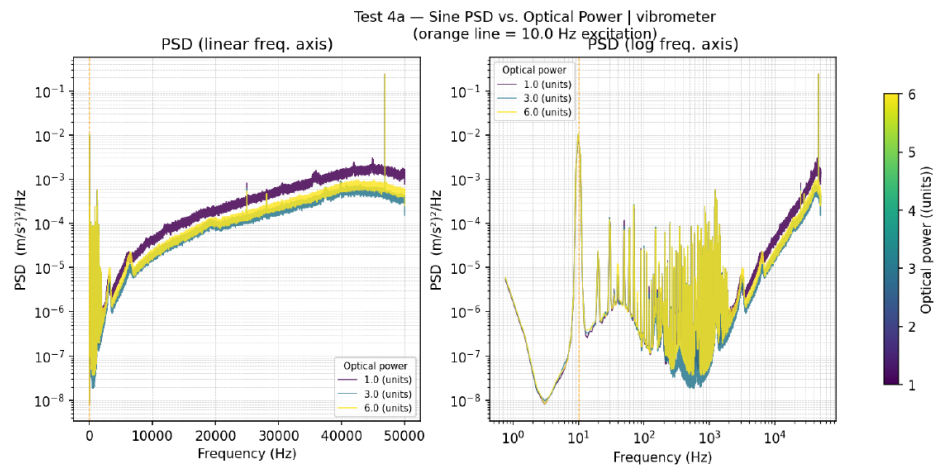


Figure A.5: T4a sine-excitation power spectral density measured by the LDV at three optical power settings, plotted against linear (left) and logarithmic (right) frequency axes, with the 10 Hz excitation marked. The measured response at the excitation frequency is unchanged across settings; optical power affects the broadband noise floor rather than the signal, and the effect is confined to frequencies well above the campaign’s measurement band.

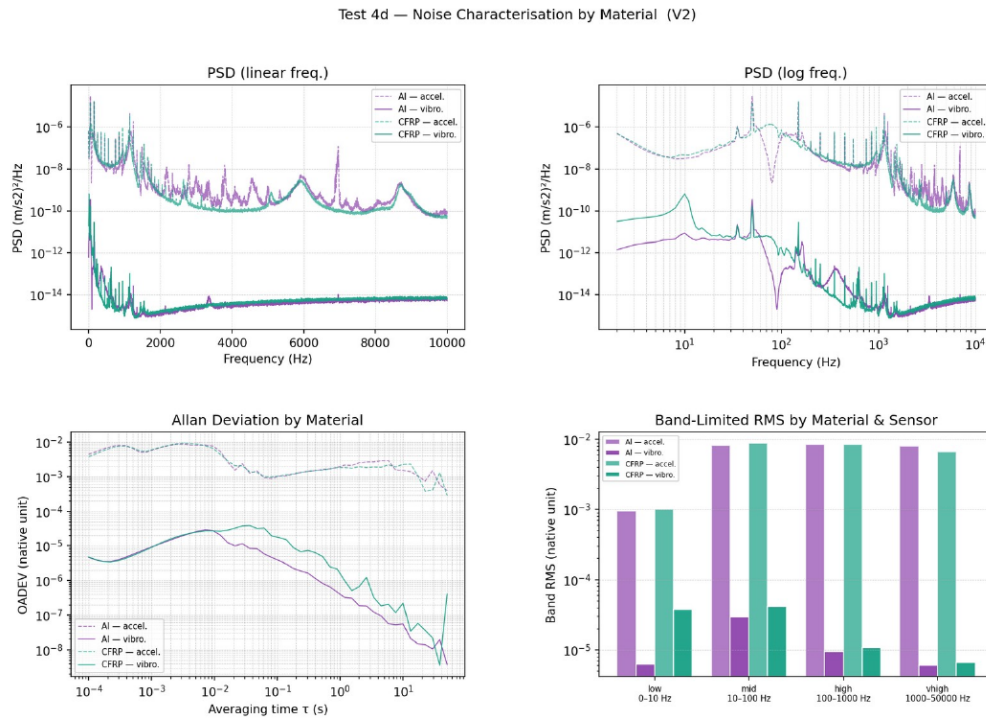


Figure A.6: T4d noise characterisation by specimen surface material, comparing aluminium against CFRP for both the accelerometer and the LDV. Clockwise from top left: noise PSD on linear and logarithmic frequency axes, band-limited RMS by frequency band, and overlapping Allan deviation against averaging time. The LDV's noise floor on the aluminium surface is roughly two orders of magnitude below the accelerometer's across every band, which is the basis for the decision not to apply a surface treatment to the specimens (Section 3.4.1).

The stand-off distance (T4b) and beam alignment (T4c) sub-tests were run over the same configuration range and produced no measurable change in noise floor or SNR within the working envelope used for the campaign; no figure is reproduced for them.

B

Modal Identification Method Selection

Section 6.1.1 of Chapter 6 reports that the conventional frequency-domain identification method produced ambiguous stabilisation diagrams for this campaign's measurements, and that a time-domain subspace method was adopted in its place. This appendix sets out why, since the reasoning turns on the interaction between two specific measurement conditions and the internal mechanics of each algorithm rather than on any property of the specimens.

Of the five modal identification methods available in the pipeline (`circle_fit`, N4SID, RFP, LSCF, and SSI-COV – of which `circle_fit`, LSCF, and N4SID are surveyed in Section 2.6.1 of Chapter 2), LSCF – the frequency-domain, matrix-fraction method that is the conventional industrial standard – was found in practice to produce ambiguous stabilisation diagrams for this campaign's baseline FRF measurements, with genuine structural poles and spurious noise-driven poles not cleanly separable across the tested model-order range. This is the expected failure mode for a frequency-domain global fit under exactly these two conditions: LSCF fits a single rational polynomial across the full measured band, so broadband noise at every frequency pulls the fit away from the true poles simultaneously, rather than being locally contained the way it would be for a per-mode method; and when two modes are closely spaced or heavily damped, their resonance peaks overlap in the frequency domain before the polynomial fit ever has a chance to separate them, since the fit's ability to resolve two poles depends on their separation relative to their combined bandwidth, which shrinks precisely as damping increases. N4SID, the time-domain subspace alternative, does not share this failure mode for two structural reasons: it estimates a discrete-time state-space realisation directly from the raw time samples via a Hankel-matrix singular value decomposition, so it never requires a clean FRF peak to already exist as a precondition for fitting; and its model order is selected from the decay of the singular value spectrum rather than from visual or automated pole-clustering across a frequency-domain sweep, which remains meaningful even when two modes' frequency responses visually overlap, provided the underlying state-space realisation still has separable eigenvalues. This is consistent with Chapter 4's own account of why N4SID was adopted as the working method for this thesis (Section 4.1.3), and with Volume 1 of the code documentation, which independently identifies N4SID as the recommended remedy specifically when the default circle-fit method produces a high rate of non-convergent (NaN) damping estimates under short or noisy scan acquisitions – exactly the symptom reported for several blocks in this campaign (Section 5.2).

The practical consequence for this thesis is recorded in Section 4.1.3 of Chapter 4, where N4SID is specified as the working identification method, and in Section 6.1.1, where the residual pole fluctuation remaining even under N4SID is discussed.

C

Supplementary Results Figures

This appendix collects figures referenced from Chapter 5 that supplement, rather than carry, the results reported there. The remaining-life panels of Section C.1 share the layout of Figures 5.19–5.21 in the main text and are reproduced here for the three formulations discussed more briefly; the fractography figures of Section C.2 provide supporting detail for Section 5.7.2.

C.1. Remaining Useful Life Panels: Remaining Formulations

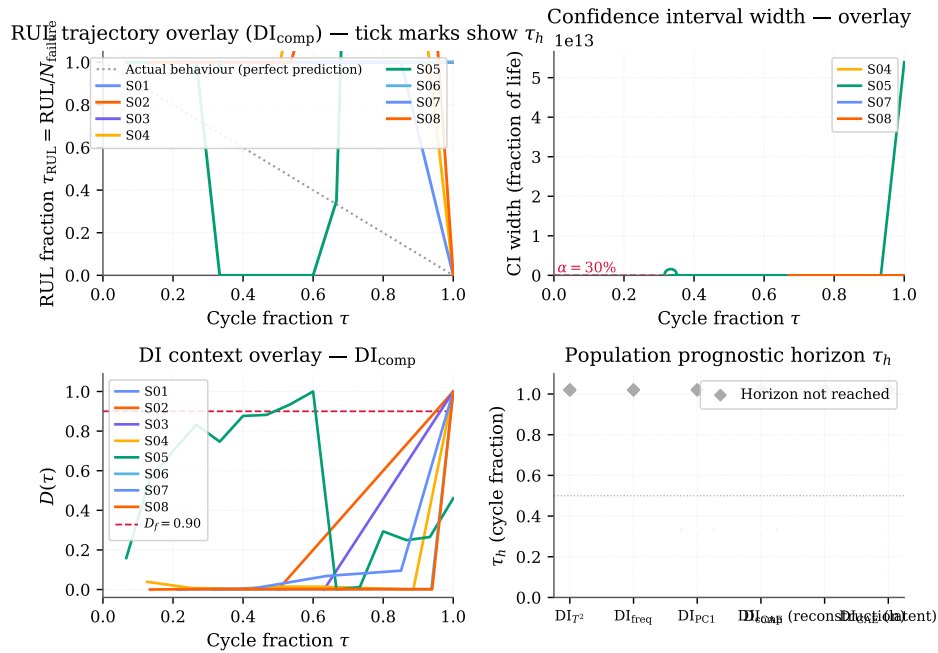


Figure C.1: RUL trajectory (Layer 3 native export), D_{comp} formulation, all eight Group C1 specimens overlaid. Panel layout identical to Figure 5.19.

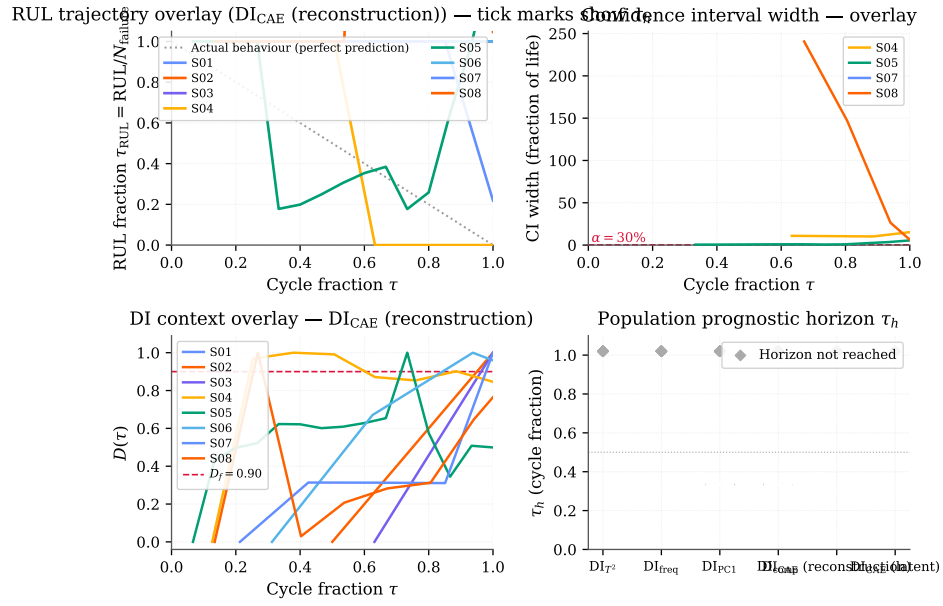


Figure C.2: RUL trajectory (Layer 3 native export), D_{CAE} (reconstruction) formulation, all eight Group C1 specimens overlaid. Panel layout identical to Figure 5.19.

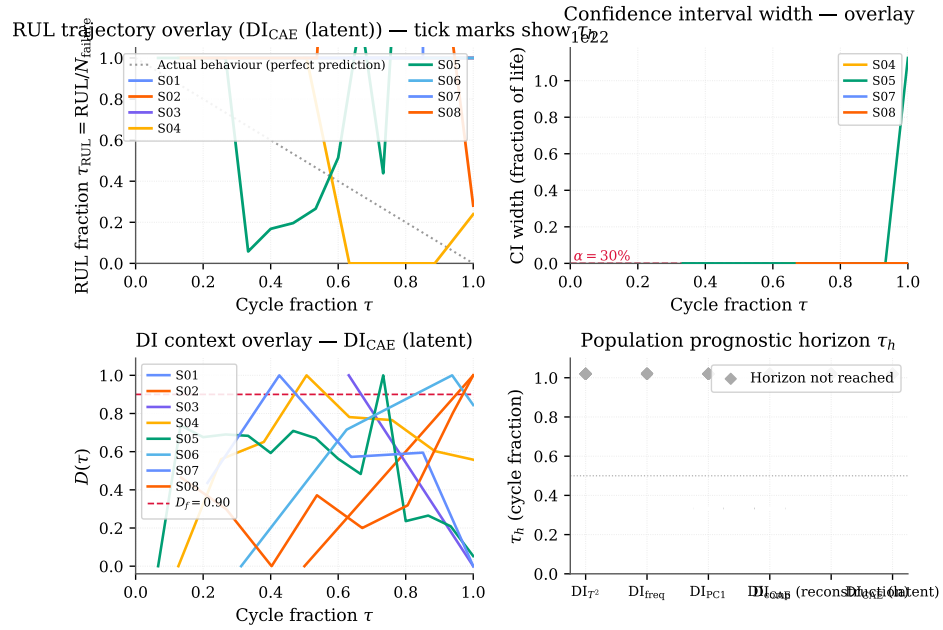


Figure C.3: RUL trajectory (Layer 3 native export), $D_{CAE,latent}$ formulation, all eight Group C1 specimens overlaid. Panel layout identical to Figure 5.19.

C.2. Supplementary Optical Fractography

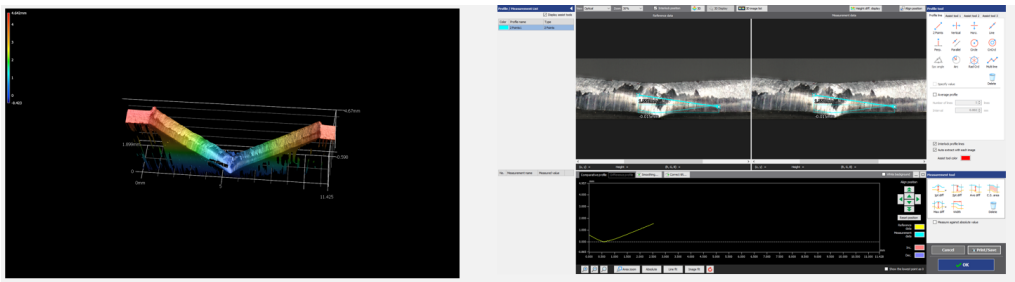


Figure C.4: Specimen C2-01 – Stage 1 pre-test surface-roughness characterisation of the notch root (profilometer export). Reported values: S_a (arithmetic mean height) and S_z (maximum height) across sampled areas at the notch root, together with the 3D height-mapped surface. This is the undamaged reference against which the Stage 3 comparison (Figure 5.26) is read.

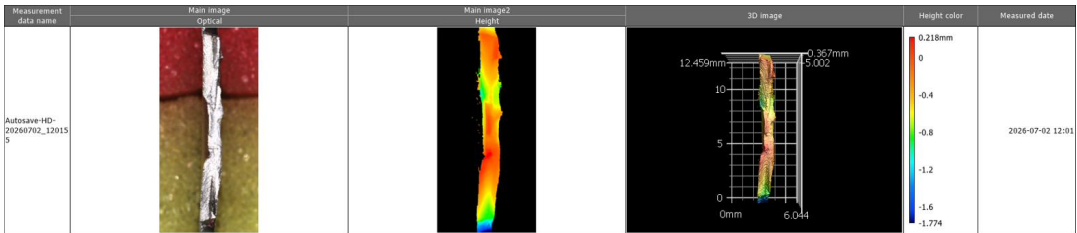


Figure C.5: Specimen C2-01 – Stage 4 fracture surface, formal profilometer export (optical, height-mapped, and 3D reconstruction views).

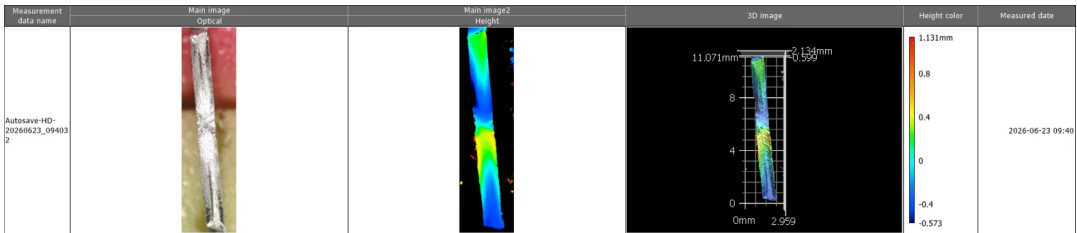


Figure C.6: Specimen C2-02 – Stage 4 fracture surface, formal profilometer export (optical, height-mapped, and 3D reconstruction views).

D

Software Toolkit and Code Availability

Three software tools were developed for this thesis. Each is released publicly in its own repository, together with the technical documentation summarised below, so that any result reported in this thesis can be traced to the code that produced it and reproduced from the same inputs.

Tool	Repository
LDV Analysis Suite	https://github.com/Anfoso/The-LDV-Analysis-Suite
Synthetic Data Toolkit	https://github.com/Anfoso/Synthetic-Data-Toolkit
Fatigue Life Pipeline	https://github.com/Anfoso/The-Fatigue-Life-Pipeline

D.1. The Three Tools

The LDV Analysis Suite performs the pre-campaign characterisation described in Section 3.4. It implements the T1–T4 sensor characterisation battery (Section 3.4.1), which establishes the noise floor, minimum detectable velocity, usable amplitude range, and disturbance sensitivities of the specific vibrometer installation, and the S1–S5 sample characterisation battery (Section 3.4.2), which establishes each specimen’s modal parameters, measurement uncertainty, and linear excitation range before fatigue loading begins. It is also the tool used to analyse the Group D1 diagnostic specimens (Section 5.8.1).

The Synthetic Data Toolkit implements the physics-informed forward model of Section 3.1: a coupled Basquin–continuum damage mechanics formulation (Section 3.1.1) evaluated at three levels of geometric fidelity (Section 3.1.2), from a closed-form analytical beam through to a meshed finite element model that resolves the notch as an explicit feature. It generates synthetic modal records with configurable sensor noise (Section 3.1.3), which were used to select the specimen geometry, bound the expected indicator response, and provide the model prediction against which the experimental record is compared in Section 6.1.6. Tier 3 requires a finite element solver; Tiers 1 and 2 do not.

The Fatigue Life Pipeline is the analysis chain of Chapter 4, comprising the five layers, the S-N characterisation module, and the live-monitoring path. It reads the HDF5 block records written during testing (Section 3.5.4) and produces every result reported in Chapter 5. Its layers are independently runnable, so an individual stage can be re-executed or replaced without repeating the whole chain.

D.2. Traceability: Thesis Section to Software Module

The table below maps each analysis stage described in this thesis to the pipeline component that implements it, and to the section reporting its output.

Table D.1: Mapping between analysis stages, the thesis sections that specify and report them, and the corresponding pipeline components.

Stage		Method	Results	Component
Feature (Layer 0)	extraction	§4.1, §4.1.6	§5.2	Baseline and fatigue block processing; shared feature extraction
Anomaly (Layer 0)	detection	§4.1.7	§5.2	Hotelling T^2 and CUSUM
FRF (Layer 0.5)	autoencoder	§4.2	§5.3	Convolutional autoencoder
Indicator (Layer 1)	ranking	§4.3	§5.3	Correlation, PCA, and SHAP-based sensitivity ranking
Damage indices (Layer 2)		§4.3.4	§5.4	Six DI formulations; logistic S-curve fitting
Remaining life (Layer 3)		§4.4	§5.5	Trend extrapolation; Gaussian Process population model; prognostic horizon
S-N characterisation		§4.5	§5.6	Force characterisation, Basquin fit, lead-time overlay
Live monitoring		§4.6	§5.7.1	Block polling and alarm evaluation

D.3. Documentation and Reproduction

Each tool is accompanied by its own technical documentation, written independently of this thesis and intended for a reader who wishes to extend or re-run the analysis rather than only to read its results. The Fatigue Life Pipeline is documented in three volumes covering, respectively, the theory and equations behind each layer, the software architecture and the design decisions behind it, and a step-by-step user guide. The Synthetic Data Toolkit is documented in two volumes covering theory and usage. The LDV Analysis Suite has a single user manual.

The documentation records known issues and unresolved limitations alongside the working functionality, including bugs found and fixed during development. This is deliberate: several of the limitations discussed in Section 7.4 are properties of the implementation rather than of the method, and separating the two requires the implementation to be inspectable.

All three tools are written in Python and depend only on widely available scientific packages. Exact version requirements, installation steps, and the expected input data layout are given in the accompanying documentation rather than reproduced here, since they are properties of the software rather than findings of this research.

Declaration on the Use of Generative AI

I acknowledge the use of generative AI tools in the preparation of this thesis and of the accompanying software. The uses are set out below by category.

Tools used

The tools used were Claude models developed by Anthropic, accessed through the Claude web interface. Several versions were used over the course of the project as they became available, principally Claude Sonnet 4.5, Claude Opus 5 and Claude Fable 5.

Refining language

Generative AI was used to improve the clarity, grammar and readability of text I had already drafted, and to identify repetitive or imprecise passages. The substance of every argument, the interpretation of every result, and the conclusions drawn from them are my own, and all suggested wording was reviewed and accepted, modified or rejected by me.

Assisting with programming and debugging

Generative AI was used substantially in developing the three software tools described in Appendix D, in three ways: the graphical user interfaces of all three were largely AI-generated; individual functions were generated from a specification I provided, stating the inputs, required behaviour and expected outputs; and AI was used for debugging and for restructuring existing code into the modular layer architecture of Chapter 4.

The analysis methods themselves are my own: which indicators to compute and how to combine them, the mathematical formulation of each layer, the selection of algorithms, and the parameter values used throughout. Generative AI implemented these decisions in code; it did not make them. All generated code was tested against expected behaviour before use, and I am able to explain and reproduce the operation of every component that produced a result reported in this thesis.

Editorial revision of the manuscript

Generative AI was used as a reviewing aid during revision of the manuscript. This went beyond language refinement and is declared separately for that reason. It was used to check cross-references and numerical consistency between chapters, to identify repeated material, to flag statements not supported by the results as reported, and to suggest reorganisation of material between sections. In places it drafted passages that I then revised, and proposed readings of experimental figures. Every factual claim, figure reading and numerical value was verified by me against the source data before any revision was accepted.

Statement of responsibility

I am personally and fully responsible for the content of this thesis, including any part prepared with the assistance of generative AI. I have critically evaluated all AI-generated output and am able to explain, defend, and reproduce the substance of the work submitted. No generative AI tool is presented as an author or co-author, and the work is an accurate representation of my own intellectual contribution.

Afonso Costa
Delft, September 2026