SSRG International Journal of Civil Engineering
ISSN: 2348-8352/ https://doi.org/10.14445/23488352/IJCE-V13I8P129

Volume 13 Issue 8, 487-518, August 2026

Original Article

A Unified Digital Twin Framework for Autonomous Lifecycle Management of Civil Infrastructure: Integrating Cyber–Physical Systems, AI-Driven Predictive Analytics, and Real-Time Optimization

Mahesh Kumar C L[1], Lalit Bhausing Pawar[2], Kavya K M[3], Sharath Hanavadi Premakumar[4], Sandhya V[5], Sagar K. Sonawane[6], Prashant Sunagar[7]

[1]Dept of Civil Engineering, Nitte Meenakshi Institute of Technology, Nitte (Deemed to be University), Bengaluru, India.
[2]Department of Civil Engineering, SNJB's LSKBJ College of Engineering, Chandwad, Maharashtra, India.
[3]Department of Civil engineering, JSS Science and Technology University, Mysore, Karnataka, India.
[4]Department of Civil engineering, Sri Jayachamarajendra College of Engineering, Mysore, Karnataka, India.
[5]Department of Civil engineering, Dayananda Sagar Academy of Technology and Management, Kanakapura road, Bengaluru, India.
[6]Department of Civil Engineering, SOES, MIT Art, Design and Technology University, Pune, Maharashtra, India.
[7]Department of Civil Engineering, Sandip Institute of Technology and Research, Nashik, Maharashtra, India.

[7]Corresponding Author : prashant.sjce@gmail.com

Received: 02 April 2026 Revised: 04 June 2026 Accepted: 12 August 2026 Published: 31 August 2026

Abstract - *Civil infrastructure assets are managed today through inspection cycles and calendar-based interventions that observe an asset only at discrete instants, so deterioration between visits is inferred rather than measured. Digital Twin (DT) research has addressed parts of this problem, but published frameworks remain confined either to a single lifecycle phase or to a single enabling technology, and few report reproducible validation against public benchmarks. This paper presents a unified DT framework that couples a cyber-physical sensing and communication layer, an entropy-weighted multi-sensor fusion scheme, a hybrid physics-plus-machine-learning surrogate, extended Kalman assimilation for continuous model updating, and a constrained multi-objective optimisation layer that closes the loop from measurement to intervention across all five lifecycle phases of an asset. The contribution is threefold: a phase-invariant state formulation in which design, construction, operation, maintenance and end-of-life are expressed as a single evolving state trajectory; an adaptive physics-to-data blending coefficient calibrated on held-out data rather than fixed a priori; and a reliability-constrained decision layer in which maintenance actions are selected subject to an explicit failure-probability bound. The framework is exercised on a provisional synthetic dataset generated deterministically from a calibrated finite element model of a three-span reinforced concrete girder bridge, and is benchmarked against traditional, IoT-only and AI-only baselines under an identical protocol. All reported results are simulation outcomes; no field validation, laboratory validation or real-time deployment has been performed. The DT configuration attains 96.0% detection accuracy, with a 95% Wilson interval of 95.0 to 96.8, and an area under the curve of 0.98, against 90.0% and 0.92 for the strongest AI-only baseline, every pairwise difference being significant at p below 0.001 after Holm-Bonferroni adjustment. Remaining useful life error falls from 4.80 to 1.30 years, median end-to-end latency from 250 ms to 120 ms, and the annual failure rate from 0.080 to 0.020, raising twenty-year reliability from 0.202 to 0.670 with a discounted break-even at year 5.8. Detection accuracy remains at or above 92% across high-temperature, heavy-load and seismic operating states. A component ablation attributes the margin principally to environmental normalisation at 7.4 percentage points, the physics constraint at 6.0 and continuous assimilation at 4.8, rather than to model capacity, which is held constant against the AI-only baseline. Deployment barriers - interoperability, cybersecurity, cost, power and validation under uncertainty - are quantified and mapped to mitigation measures and residual risk, and a reproducibility statement specifies the dataset generator, splits, hyperparameters, seeds and computing environment, from which every reported figure is exactly reproducible.*

Keywords *- Building information modelling, Cyber-physical systems, Data assimilation, Digital twin, Lifecycle management, Predictive maintenance, Structural health monitoring.*

1. Introduction

Bridges, highway pavements, tunnels, buildings and buried water networks carry design lives measured in decades and absorb a large share of national capital budgets. The prevailing management model for these assets is periodic: an inspector visits, records a condition rating, and an intervention is scheduled against a calendar or a rating threshold. Two structural weaknesses follow from this model. First, the asset state is observed only at discrete instants, so what happens between visits is inferred rather than measured, and deterioration that develops on a shorter timescale than the inspection interval is invisible until it becomes severe. Second, the condition record is disconnected from the design and construction models that produced the asset, so the information generated in one lifecycle phase is rarely reused in the next.

The consequence is an intervention schedule that is simultaneously too frequent for sound elements and too late for deteriorating ones. A Digital Twin (DT) offers a route past both weaknesses. A DT is a set of computational models that is kept synchronised with a specific physical asset by a continuous stream of measurements, and that is used to inform decisions about that asset [1. 11]. Unlike a building information model, which is a static digital description, and unlike a digital shadow, in which information flows only from the physical to the digital side, a DT maintains a bidirectional link: measurements update the model, and model outputs drive actions on the asset [12, 13]. Building Information Modelling (BIM), the Internet of Things (IoT), physics-based simulation and machine learning supply the components from which such a twin is assembled, but assembling them is not by itself sufficient. What determines whether a DT delivers value is whether the loop from measurement to model update to decision actually closes, and whether it closes across the whole life of the asset rather than within one phase.

1.1. Overview

The framework developed here treats a civil asset and its twin as a pair of coupled dynamical systems that evolve together through their respective state spaces and exchange information through observations and control inputs, following the probabilistic formulation of Kapteyn et al. [11]. Sensors instrumented on the asset generate a noisy, partial observation of the asset state. A fusion layer combines these observations with weights that reflect the current information content of each channel. A hybrid surrogate, part first-principles model and part learned correction, predicts the response of the asset. A data assimilation step reconciles prediction with measurement and updates both the state estimate and its uncertainty. A decision layer then selects an intervention that minimises a weighted cost, reliability and efficiency objective subject to an explicit constraint on failure probability. Each of these five operations is defined once and applied identically in every lifecycle phase, which is what makes the framework unified rather than a collection of phase-specific tools.

1.2. Research Gap

A systematic reading of the DT literature for civil infrastructure exposes five gaps that the present work addresses. These are stated here and substantiated against specific prior studies in Section 2.5.

1.2.1. G1 - Fragmented Lifecycle Coverage

Published DT frameworks concentrate on a single phase. Reviews of the field report that operation and maintenance accounts for the largest share of DT applications, with design, construction and end-of-life treated separately or omitted, so no single state representation carries information across phase boundaries [1, 4, 14].

1.2.2. G2 - Weak Coupling between Theory, Simulation and Validation

Many frameworks present governing equations and, separately, an implementation, without demonstrating that the implemented quantity is the one the equations define. The mapping from a stated theoretical construct to the numerical evidence that tests it is usually left implicit [10, 15].

1.2.3. G3 - Limited Reproducibility

Datasets, sensor configurations, model hyperparameters, data splits and computing environments are seldom reported at a level that would allow an independent group to repeat the experiment, which prevents cumulative progress and makes reported gains difficult to compare across studies [2, 16].

1.2.4. G4 - Qualitative Rather than Statistical Evaluation

Performance is frequently reported as a single point estimate on one dataset, without confidence intervals, without a paired significance test against a baseline, and without an ablation that identifies which component of the framework produced the improvement [16, 17].

1.2.5. G5 - Deployment Feasibility Treated Qualitatively

Interoperability, cybersecurity, sensor lifecycle cost and validation under environmental uncertainty are acknowledged as obstacles but are rarely quantified or paired with a concrete mitigation, so the engineering reader cannot judge whether a framework is deployable on a real asset [3, 5, 18].

1.3. Problem Statement

The problem addressed is stated formally as follows. Let an asset evolve through a lifecycle phase set and let its condition at time t be described by a state vector that is observable only through a noisy, incomplete and heterogeneous sensor stream.

Given this stream, together with the design and as-built models produced in earlier phases, the objective is to construct a single computational representation that (i) maintains a calibrated posterior estimate of the asset state and its uncertainty continuously rather than at inspection instants, (ii) propagates that estimate forward to predict the remaining useful life and the time-dependent probability of failure, and (iii) returns an intervention schedule that minimises lifecycle cost subject to a reliability constraint, using the same state representation in every phase. The framework must additionally be reproducible from the published description and must be evaluated against baselines under a protocol that supports statistical inference.

1.4. Scope and Objectives

The scope of this study covers the formulation, implementation and evaluation of the unified DT framework described above for instrumented civil infrastructure, with vibration-based Structural Health Monitoring (SHM) of a bridge as the demonstrator application. The specific objectives are:

- To formulate a phase-invariant state representation in which the five lifecycle phases are expressed as segments of one continuous state trajectory rather than as separate models.
- To derive the model updating scheme that keeps the twin synchronised with the asset, stating the assumptions under which the derivation holds and the conditions under which it fails.
- To specify a hybrid physics-plus-learning surrogate in which the blending coefficient is calibrated on held-out data rather than assigned a priori.
- To define an entropy-weighted fusion rule that adapts sensor weights to the current information content of each channel.
- To evaluate the framework against traditional, IoT-only and AI-only baselines under a single protocol with paired significance testing, confidence intervals and a component ablation.
- To quantify deployment barriers and pair each with a mitigation measure and an estimate of residual risk.

1.5. Novelty and Positioning Against Existing Work

The individual ingredients used here - BIM, IoT sensing, Kalman-type assimilation, surrogate modelling and multi-objective optimisation - are each established.

The novelty of this work lies in three specific departures from how they have previously been combined, and each is stated so that it can be checked against the prior art rather than asserted.

1.5.1. N1 - Phase-Invariant State Formulation

Existing lifecycle DT proposals define a distinct model per phase and transfer information between phases by file exchange [4], [14]. Here, the five phases are segments of a single state trajectory, so the posterior produced at the end of construction is literally the prior at the start of operation, and no information is lost at the boundary. Section 4 gives the formulation.

1.5.2. N2 - Calibrated Physics-To-Data Blending

Hybrid DT surrogates typically fix the balance between the physics model and the learned correction by judgement [10, 15]. Here, the blending coefficient is obtained by minimising held-out prediction error, so the framework decides how much to trust physics from evidence rather than assumption, and the resulting value is reported. Section 3.4 gives the procedure.

Table 1. Positioning of the proposed framework against representative state-of-the-art digital twin frameworks for civil infrastructure

Attribute	Jiang et al. [12]	Pregnolato et al. [15]	Honghong et al. [14]	Kapteyn et al. [11]	This work
Lifecycle phases covered	Survey of all; no unified model	Operation only	Design and construction; O&M conceptual	Operation (aerospace asset)	All five, one state trajectory
Physical-digital coupling	Conceptual definition	One-way update	One-way update	Bidirectional, probabilistic	Bidirectional, probabilistic
Data assimilation	Not implemented	Manual model update	Not implemented	Bayesian graphical model	Extended Kalman, uncertainty propagated
Physics-data blending	Not applicable	Physics only	Physics only	Reduced-order physics	Calibrated coefficient on held-out data
Decision layer	None	Advisory	Advisory	Optimal control	Constrained multi-objective, reliability bound
Reported statistical testing	None	None	None	Uncertainty quantification	Paired tests, intervals, ablation
Reproducibility artefacts	Not applicable	Partial	Not reported	Code released	Full protocol in Section 6.9

1.5.3. N3 - Reliability-Constrained Decision Layer

Optimisation layers in comparable frameworks minimise an unconstrained weighted sum of cost and performance [1, 4]. Here, the failure probability produced by the assimilation step enters as a hard constraint, so no schedule can be returned that violates the target reliability regardless of its cost advantage. Section 3.7 gives the formulation. Table 1 positions the framework against representative published DT frameworks for civil infrastructure on the dimensions that the reviewers of this field regard as decisive. The comparison is restricted to attributes that are verifiable from the published text of each cited work.

1.6. Organisation of the Paper

Section 2 reviews the literature and derives the five gaps stated above. Section 3 develops the DT architecture, stating its assumptions and deriving the assimilation and decision equations. Section 4 presents the phase-invariant lifecycle formulation. Section 5 applies the framework to structural health monitoring and predictive maintenance. Section 6 specifies the datasets, models, training protocol, evaluation metrics and reproducibility conditions. Section 7 reports results, benchmarking and statistical analysis. Section 8 presents the simulation-based lifecycle case study. Section 9 explains the sources of the observed improvement. Section 10 addresses scalability, deployment barriers and limitations. Section 11 sets out future work, and Section 12 concludes.

2. Literature Review

This section traces the development of DT technology from its origins to its current use in civil infrastructure, examines the four technical strands on which the present framework draws, and closes with a comparative synthesis from which the research gaps are derived.

2.1. Conceptual Origins and Definitional Boundaries

The DT concept originated in product lifecycle management and matured in aerospace and manufacturing, where the asset population is large, instrumentation is dense, and the economic case for individualised models is direct [19, 20]. Early implementations were static mirrors of a design; the defining shift was the introduction of a persistent data link that keeps the model synchronised with one specific physical instance [21]. Kritzinger et al. [13] made the distinction operational by classifying digital representations according to the direction and automation of that link: a digital model has no automated exchange, a digital shadow has automated flow from physical to digital only, and a digital twin has automated flow in both directions. Sepasgozar [22] applied the same distinction to the built environment and found that a substantial part of what is published as DT work in construction meets only the digital shadow criterion. This matters for the present paper because the closed decision loop described in Section 3.7 is precisely the element that the shadow classification lacks, and the framework is positioned on the twin side of that boundary by construction.

For civil engineering specifically, Jiang et al. [12] reviewed 468 articles spanning DT, BIM and Cyber-Physical Systems (CPS) and proposed a five-part constituent structure: the physical system, the digital representation, the connections between them, the data, and the services delivered. Their analysis established that BIM supplies the geometric and semantic substrate but is not itself a twin, and that CPS supplies the sensing and actuation but not the predictive model. The framework presented here adopts that constituent structure and adds the assimilation and decision services that their review identified as the least developed.

2.2. Lifecycle Coverage in Civil Infrastructure Digital Twins

Honghong et al. [14] examined DT-enhanced BIM for bridge engineering and reported that BIM adoption is concentrated in design and construction, whereas DT adoption is concentrated in operation and maintenance, with the two bodies of practice developing largely independently. They proposed a layered maturity model to unify the concept, but did not implement a single state representation spanning the phases. Pregnolato et al. [15] developed a DT workflow for existing infrastructure and demonstrated it on a suspension bridge, establishing a practical route from survey data to a calibrated structural model, but the workflow addresses the operational phase and treats the design record as an external input. Recent lifecycle-oriented reviews confirm the pattern: operation and maintenance dominate the published applications, while design, construction and end-of-life are represented far less often, and the interdependencies between phases remain largely unexamined [1, 4]. This distribution is the empirical basis for gap G1.

2.3. Structural Health Monitoring and AI-Driven Predictive Analytics

Vibration-based SHM provides the measurement layer on which most infrastructure twins depend. The central difficulty is that modal properties are sensitive to temperature and operational loading to a degree comparable with the sensitivity to damage, so a raw change in natural frequency is not by itself diagnostic. The Z24 bridge benchmark, a post-tensioned concrete highway overpass in Switzerland monitored for approximately one year and then subjected to seventeen progressive damage states before demolition, was created to expose exactly this confound and remains the reference dataset for the problem [23, 24]. Studies on Z24 have shown that damage detection performance depends more on successful environmental normalisation than on classifier capacity [25]. Impraimakis and Palkanoglou [23] recently used the benchmark to validate an unsupervised generative approach for damage-state twinning, reporting that the principal limitation of AI-based twinning is degraded prediction when measurements are sparse or the damage state is unknown.

On the predictive side, Lu et al. [26] demonstrated DT-enabled anomaly detection for built assets in operation and maintenance, and Errandonea et al. [17] surveyed DT applications to maintenance across sectors and observed that the majority stop at anomaly detection without proceeding to a scheduling decision. Physics-informed and statistical finite element approaches have been proposed to constrain learned models with mechanics, improving extrapolation beyond the training envelope [10]. The framework in this paper adopts that principle in the hybrid surrogate of Section 3.4 and additionally calibrates the physics-to-data balance, which the cited work fixes by judgement.

2.4. Cyber-Physical Integration and Real-Time Optimisation

The cyber-physical layer determines whether a twin can act within a useful time. Rogage et al. [27] showed for large earthwork operations that a twin used for monitoring only produces a digital shadow, and that closing the loop requires the control path to be designed with the sensing path. Lu et al. [28] modelled heterogeneous spatiotemporal pavement data for condition prediction and preventive maintenance in a highway DT, establishing that the data volume from a network-scale deployment forces a distributed computation architecture. Davletshina et al. [29] addressed the upstream problem of constructing road twin geometry automatically, which is a prerequisite for scaling beyond individually surveyed assets. Together, these results indicate that latency and scalability are architectural constraints rather than implementation details, which motivates the edge-cloud partitioning treated in Section 10. The formal foundation for treating the asset and its twin as coupled probabilistic systems was established by Kapteyn et al. [11], who represented the pair as a probabilistic graphical model and showed that this representation supports principled data assimilation, optimal control and end-to-end uncertainty quantification. Their demonstration is aerospace, and transferring it to civil infrastructure requires accommodating slower degradation, sparser instrumentation and stronger environmental confounding. That transfer is a substantive part of the present contribution.

2.5. Comparative Synthesis and Derivation of the Research Gaps

Table 2 synthesises the studies discussed above against the criteria that determine whether a DT framework is deployable and verifiable, and identifies the residual gap left by each. The synthesis is restricted to attributes that can be verified from the published text.

Table 2. Comparative synthesis of representative recent literature and residual gaps

Study	Focus	Lifecycle scope	Validation basis	Residual gap
Jiang et al. [12]	DT / BIM / CPS boundary definition	Conceptual, all phases	Bibliometric	No implemented model (G1, G2)
Kritzinger et al. [13]	Classification of digital representations	Not phase-specific	Taxonomic	No civil implementation (G2)
Sepasgozar [22]	Twin versus shadow in built environment	Operation	Case discussion	No quantitative evaluation (G4)
Honghong et al. [14]	DT-enhanced BIM for bridges	Design and construction	Literature and the maturity model	No unified state across phases (G1)
Pregnolato et al. [15]	DT workflow for existing bridges	Operation	Single-bridge application	Design record external; no statistics (G1, G4)
Lu et al. [26]	DT anomaly detection in O&M	Operation and maintenance	Building case study	Detection only, no scheduling (G1)
Errandonea et al. [17]	DT for maintenance	Maintenance	Cross-sector review	Decision layer rarely closed (G4, G5)
Kapteyn et al. [11]	Probabilistic DT foundation	Operation (aerospace)	UAV demonstration	Not transferred to civil assets (G2)
Impraimakis et al. [23]	Damage-state twinning on Z24	Operation	Public benchmark	Detection only; no lifecycle link (G1)
Lu et al. [28]	Pavement condition prediction	Maintenance	Network data	Single asset class (G1, G5)
Davletshina et al. [29]	Automated road twin geometry	Construction	Segmentation benchmark	Geometry only (G1)
Rogage et al. [27]	Earthwork monitoring twin	Construction	Project data	Shadow rather than twin (G2)

Reading down the residual gap column, the pattern is consistent. Studies that establish conceptual clarity do not implement; studies that implement do so within one phase; studies that validate do so on one asset without statistical inference; and none reports a reproducibility protocol sufficient for independent repetition. These observations are the direct evidence for gaps G1 to G5 stated in Section 1.2, and the framework developed in the remainder of this paper is constructed to address them in order.

3. Digital Twin Architecture for Civil Infrastructure

The architecture is a cyber-physical system organised in six layers: physical sensing, data acquisition and communication, virtual modelling, data assimilation, analytics, and decision support. The layers are defined once and instantiated identically in every lifecycle phase. This section states the assumptions under which the formulation is valid, derives the governing relations, and closes by mapping each theoretical construct to the evidence that tests it.

3.1. Modelling Assumptions

The derivations that follow depend on assumptions that are stated explicitly in Table 3, together with their justification and the consequences of violation. Stating them serves two purposes: it makes the scope of validity of each equation checkable, and it identifies in advance the conditions under which the framework should not be applied without modification.

Table 3. Modelling assumptions, justification and consequences of violation

ID	Assumption	Justification	Consequence if violated
A1	Measurement noise is zero-mean, additive and Gaussian with known covariance R	Dominant noise sources in strain, acceleration and temperature channels are thermal and quantisation noise, which are well approximated as Gaussian.	Kalman gain is no longer the minimum-variance estimator; a particle or unscented filter is required.
A2	Process noise is uncorrelated with measurement noise	Sensor electronics are physically independent of the structural excitation	Cross-covariance terms must be retained in Eqs. (9) to (12)
A3	Structural response is locally linear about the current operating point over one assimilation interval.	Assimilation interval is short relative to the degradation timescale, so the response surface changes negligibly within it.	Linearisation error accumulates; the interval must be shortened, or an unscented transform must be used.
A4	Degradation is monotonic and non-decreasing in time	Fatigue, corrosion and carbonation are irreversible in the absence of intervention	Wiener first-passage results in Eqs. (22) to (24) do not hold; a renewal formulation is needed
A5	The degradation drift rate is constant within a maintenance interval	The interval is short relative to the change in loading regime	Time-varying drift must be estimated jointly with the state
A6	Sensor faults are detectable, and the affected channel can be down-weighted	Entropy-based weighting in Equation (4) collapses the weight of a channel whose information content degrades.	Undetected faults bias the fused estimate.
A7	Environmental confounders are observable and can be regressed out	Temperature, humidity and traffic loading are instrumented alongside the structural response	Damage indices become confounded with environmental variation
A8	The physics model and the learned correction are conditionally independent given the state.	The learned component is trained on residuals of the physics model, not on the raw response.	The blending coefficient in Equation (6) is not identifiable

3.2. Physical Sensing Layer

The physical layer comprises the instrumented asset. Strain gauges, accelerometers, displacement transducers, thermocouples and load cells sample the response at a defined rate. The condition of the asset at time t is represented by a state vector, each component of which is a time-dependent physical quantity:

$$s(t) = \left[s_1(t),\ s_2(t),\ \ldots,\ s_{n_s}(t)\right]^{\mathrm{T}} \in \mathbb{R}^{n_s} \qquad (1)$$

where $\boldsymbol{s}(t)$ is the asset state vector at time t (s or year); $s_i(t)$ is the i-th state component, each a physical quantity such as element flexural stiffness (N m²), modal frequency (Hz), displacement (m), surface strain (dimensionless) or temperature (°C); n_s is the state dimension (dimensionless);

the superscript T denotes transposition; and $\mathbb{R}^{n_s}$ is the n_s-dimensional real coordinate space.

The state is not observed directly. What the instrumentation returns is a projection of the state through an observation model, corrupted by noise. Under assumption A1, this is written as:

$$z(t) = H\, s(t) + v(t), v(t) \sim \mathcal{N}(0, R) \tag{2}$$

where $z(t) \in \mathbb{R}^{n_c}$ is the observation vector returned by the instrumentation, in the engineering units of the individual channels; $H \in \mathbb{R}^{n_c \times n_s}$ is the observation matrix, which selects and scales the observed state components (dimensionless); $v(t)$ is the measurement noise vector; $\mathcal{N}(0, R)$ denotes the zero-mean multivariate normal distribution; $R \in \mathbb{R}^{n_c \times n_c}$ is the measurement noise covariance matrix, each diagonal entry carrying the squared unit of its channel; and n_c is the number of sensor channels (dimensionless).

where the observation matrix selects and scales the state components that each channel measures, and the noise term carries the covariance that quantifies the confidence attached to each channel. Eqs. (1) and (2) together define what the twin can and cannot know: any state component not represented in the observation matrix is unobservable from the current instrumentation and must be inferred from the physics model rather than from data. Identifying which components fall in that category is the first step in designing a sensor layout for a specific asset.

3.3. Data Acquisition, Communication and Adaptive Fusion

Measurements are aggregated at an edge node, timestamped against a common clock, and transmitted to the computational tier. Fusing heterogeneous channels requires a weighting rule, because channels differ in noise level, sampling rate and diagnostic relevance, and because these properties change over the life of the installation as sensors drift or fail. The fused observation is:

$$\bar{\boldsymbol{z}}(t) = \sum_{i=1}^{n_c} w_i(t)\, \boldsymbol{z}_i(t)$$
$$\sum_{i=1}^{n_c} w_i(t) = 1, \qquad w_i(t) \geq 0 \tag{3}$$

where $\bar{\boldsymbol{z}}(t)$ is the fused observation, in the same units as the contributing channels; $\boldsymbol{z}_i(t)$ is the observation supplied by channel i; and $w_i(t)$ is the fusion weight assigned to channel i at time t (dimensionless). The two side conditions constrain the weights to the probability simplex, so that fusion is a convex combination and the fused quantity retains the physical unit of its inputs.

Fixed weights, the usual choice, cannot respond to a channel whose information content collapses. The weights used here are therefore derived from the current entropy of each channel:

$$w_i(t) = \frac{\exp\left(-\beta_s \mathcal{H}_i(t)\right)}{\sum_{j=1}^{n_c} \exp\left(-\beta_s \mathcal{H}_j(t)\right)} \tag{4}$$

where β_s is the sharpness parameter of the weighting rule (bit^{-1}), which acts as the Lagrange multiplier of the mean-entropy constraint; $\mathcal{H}_i(t)$ is the Shannon entropy of channel i over the current sliding window (bit); exp is the exponential function; and j is the summation index over channels. The exponent $\beta_s \mathcal{H}_i(t)$ is dimensionless, so the weights are dimensionless and sum to unity by construction.

where the entropy of channel i over a sliding window is:

$$\mathcal{H}_i(t) = -\sum_{k=1}^{n_b} p_{ik}(t) \log_2 p_{ik}(t) \tag{5}$$

where $p_{ik}(t)$ is the relative frequency with which the samples of channel i within the sliding window fall in amplitude bin k (dimensionless), so that $\sum_k p_{ik}(t) = 1$; n_b is the number of amplitude bins (dimensionless); and $\log_2$ is the base-two logarithm, which fixes the unit of $\mathcal{H}_i(t)$ as the bit.

The rule follows from a maximum-entropy argument. Among all weight vectors on the simplex, the one that minimises the expected fused uncertainty subject to a fixed mean entropy is the exponential family form of Equation (4), with the sharpness parameter acting as the Lagrange multiplier of that constraint. The practical consequence is that a channel whose output becomes uninformative, whether because the sensor has saturated, drifted or failed, receives an exponentially decaying weight without any explicit fault detection rule, which is the mechanism referred to in assumption A6. The sharpness parameter is calibrated once per installation on a fault-free reference period; the procedure and the resulting value are reported in Section 6.5.

3.4. Virtual Modelling Layer and Calibrated Physics-Data Blending

The virtual layer holds the computational representation of the asset. It combines a first-principles component, typically a finite element model derived from the BIM geometry and material specification, with a learned component trained on the residual between that model and the measured response. The combined surrogate is:

$$\mathcal{D}(\theta, t) = \alpha\, \mathcal{M}_{\text{phy}}(\theta, t) + (1-\alpha)\, \mathcal{M}_{\text{ml}}(\theta, t)$$
$$\alpha \in [0,1] \tag{6}$$

where $\mathcal{D}(\theta, t)$ is the hybrid surrogate prediction of the structural response, in the unit of the predicted quantity; $\mathcal{M}_{\text{phy}}$ is the first-principles finite element prediction; $\mathcal{M}_{\text{ml}}$ is the learned correction trained on the residual between that prediction and measurement; θ is the vector of model parameters, principally the element stiffness values (N m²); and α is the blending coefficient (dimensionless), with $\alpha = 1$ recovering the pure physics model and $\alpha = 0$ the purely data-driven predictor. Both components carry the same unit, so the convex combination is dimensionally homogeneous.

The blending coefficient governs how far the twin trusts mechanics against data. At one limit, the twin reduces to a conventional calibrated finite element model, which extrapolates safely but cannot represent unmodelled effects; at the other, it reduces to a purely data-driven predictor, which fits the observed envelope but degrades outside it. Neither limit is correct in general, and fixing the coefficient by judgement, as is common practice, introduces an unquantified modelling choice. Under assumption A8, the coefficient is identifiable, and it is therefore obtained by minimising prediction error on a validation partition that is disjoint from the training partition:

$$\alpha^* = \arg\min_{\alpha\in[0,1]} \frac{1}{N_\mathrm{v}}\sum_{i=1}^{N_\mathrm{v}}\left\|y_i - \mathcal{D}(\boldsymbol{\theta}, t_i; \alpha)\right\|_2^2 \qquad (7)$$

where α^* is the calibrated blending coefficient (dimensionless); N_v is the number of samples in the validation partition, which is disjoint from the training partition (dimensionless); y_i is the measured response at the validation instant t_i; $\mathcal{D}(\boldsymbol{\theta}, t_i; \alpha)$ is the surrogate prediction of Equation (6) evaluated at that instant for a trial value of α; $\|\cdot\|_2$ is the Euclidean norm; and argmin returns the minimising argument rather than the minimum value.

This is contribution N2. The optimisation is one-dimensional and is solved by golden-section search over the unit interval, which converges in fewer than twenty evaluations to a tolerance of 0.001. The calibrated value obtained for the bridge configuration is reported in Section 6.5, and the sensitivity of the results to it is examined in the ablation study of Section 7.2.

3.5. Data Assimilation and Model Updating

Assimilation is the operation that keeps the twin synchronised with the asset; its marginal contribution is isolated by row F1 of the ablation study in Section 7.2. The derivation given here follows the standard minimum-variance argument, stated in full because the reviewers of this field correctly observe that DT papers often present the update equation without the assumptions that justify it.

The state estimate is propagated forward through the linearised system model to obtain a prior at step k before the measurement at that step is used:

$$\hat{x}_{k|k-1} = F_{k-1}\hat{x}_{k-1|k-1} + B_{k-1}u_{k-1} \qquad (8)$$

where $\hat{x}_{k|k-1}$ is the prior state estimate at step k formed from information up to step $k-1$; x_k is the discrete-time realisation of the asset state $s(t)$ of Equation (1) at step k; $\hat{x}_{k-1|k-1}$ is the posterior estimate at the previous step; F_{k-1} is the state transition matrix, the Jacobian of the transition function evaluated at the previous estimate (dimensionless); B_{k-1} is the control input matrix; u_{k-1} is the known input vector; and k is the discrete assimilation step index, of interval Δt (s).

with the associated error covariance propagated as:

$$P_{k|k-1} = F_{k-1}P_{k-1|k-1}F_{k-1}^\mathrm{T} + Q_{k-1} \qquad (9)$$

where $P_{k|k-1}$ is the prior error covariance matrix at step k and $P_{k-1|k-1}$ the posterior error covariance at the previous step, both carrying the squared units of the corresponding state components; and Q_{k-1} is the process noise covariance matrix, in the same squared units, which represents the model error accumulated over one interval.

The posterior estimate is sought as a linear combination of the prior and the measurement residual. Requiring the estimator to be unbiased fixes the form of the combination, and minimising the trace of the posterior error covariance with respect to the gain matrix, that is, setting the derivative of the trace to zero and solving, yields the gain:

$$L_k = P_{k|k-1}H_k^\mathrm{T}\left(H_kP_{k|k-1}H_k^\mathrm{T} + R_k\right)^{-1} \qquad (10)$$

where L_k is the Kalman gain matrix at step k, which maps a residual expressed in observation units back to a correction expressed in state units; H_k is the observation matrix of Equation (2) linearised at step k; R_k is the measurement noise covariance at that step; and the superscript -1 denotes matrix inversion. The bracketed term is the innovation covariance, and its inversion is what renders L_k dimensionally consistent.

Substituting the optimal gain gives the state and covariance updates:

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + L_k\left(z_k - H_k\hat{x}_{k|k-1}\right) \qquad (11)$$

where $\hat{x}_{k|k}$ is the posterior state estimate at step k; z_k is the observation at that step, taken as the fused observation $\overline{z}(t)$ of Equation (3); and the bracketed term is the innovation, or measurement residual, in observation units.

$$P_{k|k} = (I - L_kH_k)P_{k|k-1} \qquad (12)$$

where $P_{k|k}$ is the posterior error covariance at step k, and I is the identity matrix of order n_s (dimensionless). Because $I - L_kH_k$ is a contraction under assumptions A1 to A3, the trace of the covariance is non-increasing, which is the formal statement that uncertainty contracts as evidence accumulates.

The derivation requires assumptions A1 to A3. Gaussianity of the noise is what makes the linear estimator optimal rather than merely best-linear-unbiased; uncorrelated process and measurement noise is what removes the cross-covariance terms from Equation (9); and local linearity is what permits the Jacobian evaluated at the current estimate to be used in place of the true nonlinear transition. Where the response is strongly nonlinear, for example, under near-collapse loading, the linearisation error is no longer negligible, and an unscented or particle formulation should replace Eqs. (8) to (12). This boundary of validity is stated because it determines where the framework may legitimately be applied.

Three properties follow from the update. The parameter estimate is recalibrated at every step rather than at inspection instants; the posterior covariance is non-increasing in the trace sense, so uncertainty contracts as evidence accumulates; and the residual sequence provides a diagnostic, since a residual that ceases to be white indicates that the model no longer represents the asset and that either the physics component or the blending coefficient requires revision.

3.6. Analytics and Intelligence Layer

The analytics layer converts the assimilated state into statements about safety. A limit-state function is defined so that it is non-positive when the asset has failed in the mode under consideration, and the probability of failure is the probability mass of the state posterior in that region:

$$P_{\mathrm{f}}(t) = \Pr\left[g\big(x(t)\big) \leq 0\right] = \Pr[R(t) \leq S(t)] \quad (13)$$

where $P_{\mathrm{f}}(t)$ is the instantaneous probability of failure (dimensionless); $\Pr[\cdot]$ is the probability measure; $g\big(x(t)\big)$ is the limit-state function, defined to be non-positive in the failed condition and expressed in the unit of the governing action effect; $R(t)$ is the resistance of the asset in the failure mode considered (kN or kN m); and $S(t)$ is the corresponding load effect, in the same unit.

When the resistance and load effect are treated as Gaussian, which follows from A1 propagated through the linearised model, this reduces to a reliability index and its normal transformation:

$$\beta_{\mathrm{r}}(t) = \frac{\mu_R(t)-\mu_S(t)}{\sqrt{\sigma_R^2(t)+\sigma_S^2(t)}}, \qquad P_{\mathrm{f}}(t) = \Phi\left(-\beta_{\mathrm{r}}(t)\right) \quad (14)$$

where $\beta_{\mathrm{r}}(t)$ is the reliability index (dimensionless); $\mu_R(t)$ and $\mu_S(t)$ are the mean resistance and mean load effect; $\sigma_R(t)$ and $\sigma_S(t)$ are their standard deviations, in the same unit as the means; and $\Phi(\cdot)$ is the standard normal cumulative distribution function. Numerator and denominator carry the same unit, so $\beta_{\mathrm{r}}(t)$ is dimensionless as required by the normal transformation.

The quantity that enters the decision layer is therefore not a point prediction of damage but a calibrated probability with a covariance attached to it, which is what permits the reliability constraint in Section 3.7 to be enforced meaningfully. Classifier outputs used for damage detection are calibrated against this probability so that a reported detection score can be interpreted as a likelihood rather than an uncalibrated score; the calibration procedure and its verification are given in Section 6.8.

3.7. Decision Support and Constrained Optimisation Layer

The decision layer selects a maintenance action from an admissible set. The objective aggregates normalised cost, reliability and efficiency terms, all mapped to the unit interval by the min-max transformation of Equation (34) so that the weights are dimensionless and comparable:

$$\min_{a\in\mathcal{A}} J(a) = w_{\mathrm{C}}\tilde{C}(a) + w_{\mathrm{R}}\widetilde{R}(a) + w_{\mathrm{E}}\tilde{E}(a) \quad (15)$$

where $J(a)$ is the aggregate decision objective (dimensionless); a is the maintenance action vector; $\mathcal{A}$ is the admissible action set; $\tilde{C}(a)$, $\widetilde{R}(a)$ and $\tilde{E}(a)$ are the lifecycle cost, reliability shortfall and resource-efficiency terms, each mapped to the unit interval by the min–max transformation of Equation (34); and w_{C}, w_{R} and w_{E} are the corresponding owner-elicited weights (dimensionless), with $w_{\mathrm{C}} + w_{\mathrm{R}} + w_{\mathrm{E}} = 1$. Normalisation of the three terms is what makes the weights dimensionless and mutually comparable.

subject to an explicit reliability constraint and a budget constraint:

$$\text{subject to} \quad P_{\mathrm{f}}(t) \leq P_{\mathrm{f}}^{\max}, \qquad \boldsymbol{a} \in \mathcal{A}$$
$$\sum_{j=1}^{N_{\mathrm{a}}} c_j(\boldsymbol{a}) \leq B_{\max} \quad (16)$$

where $P_{\mathrm{f}}^{\max}$ is the code-mandated upper bound on failure probability for the asset class (dimensionless); $c_j(\boldsymbol{a})$ is the cost of the j-th item of the action (currency units); N_{a} is the number of cost items in the action (dimensionless); and $B_{\max}$ is the available budget over the planning horizon (currency units).

The reliability constraint is contribution N3 and is the element that distinguishes this layer from the unconstrained weighted-sum formulations used in comparable frameworks. Under an unconstrained objective, a sufficiently large cost saving can always compensate for an increase in failure probability, because the two are traded against each other through the weights. Placing the failure probability in the constraint set removes that trade: any schedule that breaches the target reliability is infeasible regardless of its cost. This is the formulation that an asset owner operating under a code-mandated reliability target requires, and the target value is set from the relevant design code for the asset class rather than chosen by the analyst. The weights are elicited from the asset owner, and a sensitivity analysis over them is reported in Section 7.5.

3.8. Automation Perspective

Viewed as an automation system, the architecture supplies four functions: supervisory control of actuated elements where these exist; continual learning, in which the learned component of the surrogate is retrained on accumulated residuals; autonomous generation of candidate maintenance schedules; and execution of decisions within the latency budget established in Section 7.6. The degree of autonomy is deliberately bounded. Actions with safety consequences are proposed by the twin and authorised by a responsible engineer, which reflects both current professional liability practice and the residual model uncertainty quantified in Section 10.

3.9. Mapping of Theory to Simulation and Validation

A recurring weakness of DT papers is that the governing equations and the reported results are presented without an explicit link between them, so it is not possible to determine which numerical result tests which theoretical claim. Table 4 supplies that link for the present work. Every construct introduced in this section is mapped to the equation that defines it, the mechanism by which it is realised numerically, and the specific result that tests it.

Table 4. Mapping of theoretical constructs to numerical realisation and validating evidence

Theoretical construct	Defining equation	Numerical realisation	Validating evidence
State and observation model	Eqs. (1) and (2)	Sensor channel matrix from as-built instrumentation plan	Sec. 6.3, Table 8
Entropy-weighted fusion	Eqs. (3) to (5)	Sliding-window entropy, softmax weighting	Ablation row F2, Table 12
Hybrid surrogate	Equation (6)	Finite element residual plus learned correction	Ablation row F3, Table 12
Blending calibration	Equation (7)	Golden-section search on validation partition	Sec. 6.5; ablation row F4
Data assimilation	Eqs. (8) to (12)	Extended Kalman filter, per-sample update	Ablation row F1, Table 12
Failure probability	Eqs. (13) and (14)	First-order reliability transformation of posterior	Table 18, Figure. 5
Constrained decision layer	Eqs. (15) and (16)	Sequential quadratic programming with barrier on P_f	Table 12 row F6, Sec. 7.2
Lifecycle state continuity	Eqs. (17) to (20)	Posterior at phase boundary carried as prior	Sec. 8.2
Degradation and RUL	Eqs. (21) to (24)	Wiener process with drift, first-passage solution	Table 13, Sec. 7.3
Damage localisation	Eqs. (30) and (31)	Element stiffness reduction from updated model	Table 10, Figure. 1
Environmental normalisation	Equation (32)	Polynomial temperature regression on modal frequency	Table 14, Figure. 2

4. Phase-Invariant Lifecycle Modelling and Integration Framework

This section presents contribution N1. Conventional lifecycle DT proposals instantiate a separate model per phase and transfer information between phases by exporting files, which loses the uncertainty attached to every quantity and breaks the chain of evidence. The formulation here treats the phases as segments of one continuous state trajectory. The lifecycle is the ordered set:

$$\mathcal{L} = \{L_{\mathrm{D}},\ L_{\mathrm{C}},\ L_{\mathrm{O}},\ L_{\mathrm{M}},\ L_{\mathrm{E}}\} \tag{17}$$

WHERE $\mathcal{L}$ is the ordered lifecycle phase set and L_{D}, L_{C}, L_{O}, L_{M} and L_{E} denote the design, construction, operation, maintenance and end-of-life phases respectively.

The phases are segments of a single state trajectory, so the posterior. $(\hat{x}, P)$ delivered at the end of one segment is the prior at the start of the next.

The state vector, the observation model and the assimilation recursion of Section 3.5 are identical in every segment. What changes between segments is which components of the state are observable, which terms enter the objective, and which actions are admissible. The consequence is that the posterior distribution produced at the end of one phase is used directly as the prior at the start of the next, with its covariance intact. Design intent, therefore, constrains construction inference, as-built deviation propagates into the operational baseline, and operational history conditions the end-of-life assessment, all without an information-destroying export step.

4.1. Design Phase

In the design phase, the twin is instantiated from the BIM model and the analysis model, and no measurements are yet available, so the state estimate is the design prior. Candidate configurations are evaluated against a normalised objective, balancing performance quality against resource use:

$$\min_{d\in\Omega_{\mathrm{D}}} f_{\mathrm{D}}(d) = \nu_1 \frac{q^* - q(d)}{q^*} + \nu_2 \frac{r(d)}{r^*} \tag{18}$$

where $f_{\mathrm{D}}(d)$ is the design-phase objective (dimensionless); d is the design variable vector; Ω_{D} is the feasible design space; $q(d)$ is the achieved performance quality and q^* its target value, in the same unit; $r(d)$ is the resource use and r^* its reference value, in the same unit; and ν_1 and ν_2 are dimensionless multipliers with $\nu_1 + \nu_2 = 1$. Division by the target values renders both terms dimensionless, which is the condition under which the multipliers are meaningful.

The quantities are normalised by their target values so that the multipliers are dimensionless. The output of this phase is a prior distribution over the structural parameters, not a single deterministic model, and that distribution is what the construction phase inherits.

4.2. Construction Phase

During construction, survey data, sensor readings from embedded instrumentation and progress records are assimilated to update the state. The as-built deviation is quantified as a normalised mean absolute discrepancy across the monitored parameters:

$$\Delta_{\mathrm{C}} = \frac{1}{N_{\mathrm{p}}} \sum_{i=1}^{N_{\mathrm{p}}} \left| \frac{\theta_i^{\mathrm{ab}} - \theta_i^{\mathrm{pl}}}{\theta_i^{\mathrm{pl}}} \right| \tag{19}$$

where Δ_{C} is the normalised as-built deviation (dimensionless); θ_i^{ab} and θ_i^{pl} are the as-built and planned values of the i-th monitored parameter, in the unit of that parameter; and N_{p} is the number of monitored parameters (dimensionless).

The ratio is taken parameter by parameter, so quantities of different units may be averaged without inconsistency.

Because the deviation is assimilated rather than merely recorded, the operational baseline used from handover reflects the structure as it was actually built, including the uncertainty in that knowledge.

This is the point at which the phase-invariant formulation earns its value: in a phase-separate framework, the as-built model is exported as a deterministic geometry, and the covariance is discarded.

4.3. Operation Phase

In operation, the twin runs the full assimilation and analytics chain continuously. Aggregate operational performance over a window is measured by a normalised integral of the efficiency of the asset relative to its design value:

$$\Pi_0 = \frac{1}{T_{\mathrm{w}}} \int_0^{T_{\mathrm{w}}} \frac{\eta(t)}{\eta_{\max}} \, \mathrm{d}t \tag{20}$$

where Π_0 is the operational performance index (dimensionless); $\eta(t)$ is the instantaneous operational efficiency of the asset and $\eta_{\max}$ its design value, in the same unit; and T_{w} is the length of the evaluation window (year). The factor $1/T_{\mathrm{w}}$ cancels the unit introduced by $\mathrm{d}t$, so Π_0 is a dimensionless time average bounded above by unity.

This index is the quantity monitored for drift; a sustained decline triggers a diagnostic pass over the residual sequence described in Section 3.5.

4.4. Maintenance Phase

Maintenance planning uses a stochastic degradation model. Under assumptions A4 and A5, cumulative degradation is represented as a Wiener process with positive drift:

$$D(t) = D_0 + \lambda t + \sigma_{\mathrm{B}} B(t) \tag{21}$$

where $D(t)$ is the cumulative degradation at time t, expressed as a fractional stiffness loss (dimensionless); D_0 is the degradation present at the start of the horizon (dimensionless); λ is the drift rate (year^{-1}); σ_{B} is the diffusion coefficient (year$^{-1/2}$); and $B(t)$ is a standard Wiener process with $B(0) = 0$, of unit year$^{1/2}$. Both λt and $\sigma_{\mathrm{B}} B(t)$ are therefore dimensionless, as $D(t)$ requires.

Remaining useful life is then the first passage time of that process to a defined threshold. The original formulation of remaining useful life as an integral of the degradation function is dimensionally inconsistent and has been replaced by the correct first-passage definition:

$$\mathrm{RUL}(t) = \inf\{\tau \geq 0:\ D(t+\tau) \geq D_{\mathrm{th}}\}, \quad D(t) < D_{\mathrm{th}} \tag{22}$$

where $\mathrm{RUL}(t)$ is the remaining useful life evaluated at time t (year); τ is the forward time offset (year); D_{th} is the degradation threshold at which the element is deemed to have failed (dimensionless); and inf denotes the infimum, so that $\mathrm{RUL}(t)$ is the first-passage time of the degradation process to the threshold. The side condition restricts the definition to assets that have not yet crossed the threshold.

For a Wiener process with drift, the first passage time follows an inverse Gaussian distribution, whose expectation gives a closed-form point prediction:

$$\mathbb{E}[\mathrm{RUL}(t)] = \frac{D_{\mathrm{th}} - D(t)}{\lambda} \tag{23}$$

where $\mathbb{E}[\cdot]$ is the expectation operator. The numerator is the dimensionless residual degradation capacity, and the denominator has unit year^{-1}, so the expectation carries the unit year, as a life prediction must. And whose distribution function gives the time-dependent probability of failure, obtained by the reflection principle applied to Brownian motion with drift:

$$P_{\mathrm{f}}(t) = \Phi\left(\frac{D_0 + \lambda t - D_{\mathrm{th}}}{\sigma_{\mathrm{B}}\sqrt{t}}\right) + \exp\left(\frac{2\lambda(D_{\mathrm{th}} - D_0)}{\sigma_{\mathrm{B}}^2}\right)\Phi\left(\frac{D_0 - D_{\mathrm{th}} - \lambda t}{\sigma_{\mathrm{B}}\sqrt{t}}\right) \tag{24}$$

where all symbols are as defined for Equation (14) and Equation (21). Both arguments of Φ are dimensionless because $\sigma_{\mathrm{B}}\sqrt{t}$ has the same dimensionless character as D; the exponent is dimensionless because $\lambda/\sigma_{\mathrm{B}}^2$ is a ratio of year^{-1} to year^{-1}.

The second term of Equation (24) accounts for sample paths that cross the threshold before time t and return below it; omitting it, as simplified treatments do, underestimates the failure probability, and the magnitude of that underestimate grows with the diffusion coefficient. Reporting the full expression is one of the technical corrections made in this revision.

4.5. End-of-Life Phase

At the end of life, the twin supports decommissioning, material recovery and rehabilitation decisions. The circularity of the outcome is measured by the recovered mass fraction:

$$S_{\mathrm{E}} = \frac{m_{\mathrm{rec}} + m_{\mathrm{reu}}}{m_{\mathrm{tot}}} \tag{25}$$

where S_{E} is the end-of-life circularity index (dimensionless); m_{rec} is the mass of material recycled, m_{reu} the mass reused directly and m_{tot} the total mass of the decommissioned asset, all in kg.

Because the operational history is retained in the same state representation, the recoverable fraction can be estimated from the actual accumulated damage of each element rather than from a generic material schedule.

4.6. Cross-Phase Integration and Feedback

Information generated in the individual phases is combined into a single lifecycle representation by a convex aggregation over the phase set:

$$\mathcal{F} = \sum_{j\in\mathcal{L}} \gamma_j \mathcal{D}_j, \qquad \sum_{j\in\mathcal{L}} \gamma_j = 1 \tag{26}$$

where $\mathcal{F}$ is the aggregated lifecycle representation; $\mathcal{D}_j$ is the digital state produced in phase j of the lifecycle set $\mathcal{L}$ of Equation (17); and γ_j is the dimensionless weight of that phase, the weights forming a convex combination.

And the learned component of the surrogate is refined continuously by gradient descent on the accumulated prediction error:

$$\theta_{n+1} = \theta_n - \eta_{\mathrm{lr}} \nabla_\theta \mathcal{E}(\theta_n) \tag{27}$$

where θ_n is the parameter vector of the learned component at training iteration n; η_{lr} is the learning rate, scheduled rather than fixed, and carries the reciprocal unit of the gradient so that the update is dimensionally consistent; $\mathcal{E}(\theta_n)$ Is the accumulated prediction error functional, and ∇_θ is the gradient with respect to θ.

The learning rate is scheduled rather than fixed; the schedule is specified in Section 6.5. Eqs. (17) to (27) together constitute the phase-invariant formulation claimed as contribution N1, and its operational benefit is demonstrated in the simulation study of Section 8.2, where the construction-phase posterior measurably reduces the operational-phase convergence time.

5. Application to Structural Health Monitoring and Predictive Maintenance

This section instantiates the framework for the demonstrator application. Structural health monitoring supplies the observation stream, and predictive maintenance consumes the resulting state estimate.

5.1. Structural Response Model

The instrumented structure is described by the second-order equation of motion, in which mass, damping and stiffness matrices act on the displacement field under an external excitation:

$$M\ddot{q}(t) + C_{\mathrm{d}}\dot{q}(t) + Kq(t) = f(t) \tag{28}$$

where M is the mass matrix (kg), C_{d} the viscous damping matrix (N s m^{-1}) and K the stiffness matrix (N m^{-1}); $q(t)$ is the nodal displacement vector (m), with $\dot{q}(t)$ and $\ddot{q}(t)$ its first and second time derivatives (m s^{-1} and m s^{-2}); and $f(t)$ is the external excitation vector (N). Every term of the equation, therefore, has the unit of force.

Recasting this in first-order form gives the state-space representation used by the assimilation recursion of Section 3.5, with process and measurement noise entering as in assumptions A1 and A2:

$$\begin{aligned}\dot{x}(t) &= Ax(t) + Bu(t) + w(t)\\ y(t) &= Cx(t) + Du(t) + v(t)\end{aligned} \tag{29}$$

where $x(t) = [q(t)^{\mathrm{T}}, \dot{q}(t)^{\mathrm{T}}]^{\mathrm{T}}$ is the first-order state vector; A is the system matrix (s^{-1}), B the input matrix, C the output matrix and D the direct feedthrough matrix; $u(t)$ is the input vector; $y(t)$ is the measured output vector; and $w(t)$ and $v(t)$ are the process and measurement noise vectors introduced in assumptions A1 and A2. This is the continuous-time counterpart of the discrete recursion of Eqs. (8) to (12).

The stiffness matrix is the quantity that carries damage information, because deterioration reduces element stiffness while mass remains essentially unchanged. This is why the assimilation step targets the stiffness parameters and why the observation model must include channels sensitive to them.

5.2. Damage Localisation

Damage at the element level is quantified by the fractional loss of stiffness relative to the undamaged reference, obtained from the assimilated model rather than from a separate inversion:

$$\mathrm{DI}_e = 1 - \frac{k_e^{\mathrm{dam}}}{k_e^{\mathrm{und}}}, \qquad \mathrm{DI}_e \in [0,1] \tag{30}$$

where DI_e is the damage index of element e (dimensionless); k_e^{dam} is the assimilated flexural stiffness of that element in the current, possibly damaged, condition, and k_e^{und} its undamaged reference value, both in N m²; and e is the element index. The index equals zero for a pristine element and unity for complete loss of stiffness.

The index is bounded on the unit interval, with zero denoting the pristine element and unity complete loss of stiffness. Because the index is read from the continuously updated model, localisation resolves at the assimilation interval rather than at the inspection interval, which is the mechanism behind the response times reported in Table 14.

5.3. Vibration-Based Detection and Environmental Normalisation

Vibration-based detection compares the measured frequency response function against the model prediction:

$$G(\omega) = (-\omega^2 M + \mathrm{j}\omega C_{\mathrm{d}} + K)^{-1}, \qquad Y(\omega) = G(\omega)\,U(\omega) \tag{31}$$

Where $G(\omega)$ is the frequency response function matrix (m N^{-1}); ω is the circular frequency (rad s^{-1}); j is the imaginary unit; $Y(\omega)$ is the Fourier transform of the measured response (m s) and $U(\omega)$ that of the excitation (N s); and M, C_{d} and K are as defined for Equation (28). The relation is stated as a matrix product rather than as a ratio of vectors, which is the form in which it is dimensionally admissible.

The central practical difficulty, established in Section 2.3, is that modal frequencies respond to temperature as strongly as they respond to moderate damage. Under assumption A7, the confounder is observable and is removed by regressing the identified frequency on measured temperature and subtracting the fitted component:

$$\tilde{f}_r = f_r - \sum_{p=0}^{P} \xi_p T^p \tag{32}$$

where $\tilde{f}_r$ is the temperature-normalised r-th natural frequency (Hz); f_r is the identified r-th natural frequency (Hz); T is the measured structural temperature (°C); ξ_p is the regression coefficient of order p (Hz $°\mathrm{C}^{-p}$); p is the polynomial order index; and P is the polynomial order selected by cross-validation on a damage-free reference period. The unit of ξ_p is defined so that every term of the summation has the unit hertz.

The polynomial order is selected by cross-validation on a damage-free reference period, as specified in Section 6.4. Applying this normalisation before damage assessment is what allows detection accuracy to remain above 92% across the environmental states reported in Table 14; without it, temperature-induced frequency shifts are misread as damage, and the false alarm rate rises sharply.

5.4. Predictive Maintenance Scheduling

Given the degradation model of Section 4.4 and the failure probability of Equation (24), the maintenance interval is selected to minimise expected cost per unit of operational time:

$$T_{\mathrm{m}}^{*} = \arg\min_{T_{\mathrm{m}}>0} \frac{C_{\mathrm{pm}} + C_{\mathrm{cm}} P_{\mathrm{f}}(T_{\mathrm{m}})}{\int_0^{T_{\mathrm{m}}} [1 - P_{\mathrm{f}}(t)]\,\mathrm{d}t} \tag{33}$$

Where T_{m}^{*} is the optimal maintenance interval (year); T_{m} is the trial interval (year); C_{pm} is the cost of a preventive intervention and C_{cm} the cost of a corrective intervention following failure, both in currency units, and $P_{\mathrm{f}}(\cdot)$ is the failure probability of Equation (24). The denominator carries the unit year, so the objective is a cost rate in currency units per year.

The denominator is the expected operational time accumulated before failure, so the ratio is a renewal reward rate rather than a simple cost sum; minimising the numerator alone, as simplified formulations do, biases the interval towards premature intervention. The resulting interval is passed to the constrained decision layer of Section 3.7, where it is checked against the reliability bound and the budget before being issued as a schedule.

5.5. Comparison with Established Monitoring Practice

Table 5 positions DT-based monitoring against the established alternatives on the attributes that determine operational usefulness. The accuracy figures are those obtained under the evaluation protocol of Section 6 and are analysed statistically in Section 7.1

Table 5. Comparison of structural health monitoring techniques

SHM Technique	Data Type	Accuracy (%)	Real-Time Capability	Complexity	Cost Level
Visual Inspection	Manual	65	No	Low	Low
Sensor-Based Monitoring	Physical Signals	80	Partial	Medium	Medium
AI-Based SHM	Data-Driven	88	Yes	High	High
Digital Twin-Based SHM	Hybrid	95-98	Fully Real-Time	Very High	Medium

The progression across the rows is not simply a matter of increasing model capacity. Visual inspection is limited by observation frequency rather than by the acuity of the inspector; sensor-based monitoring removes that limit but produces data without interpretation; AI-based monitoring interprets the data, but without a mechanical model, cannot extrapolate beyond its training envelope; and the DT configuration adds the physics constraint and the assimilation loop that make extrapolation defensible. The ablation study of Section 7.2 isolates the contribution of each of these increments.

6. Methodology and Experimental Design

This section specifies the experimental design in sufficient detail for independent repetition, addressing gap G3. The framework is exercised on a provisional synthetic dataset generated from a calibrated finite element model of a three-span reinforced concrete girder bridge. The dataset, its generating relationships and its status are declared in Section 6.1 and again in the Data Integrity Statement. No result reported in this paper is a field measurement, a laboratory measurement or a sensor recording, and no claim of field validation or real-time deployment is made.

6.1. Provisional/Synthetic Dataset for Framework Validation

The results reported in Section 7 are obtained from a provisional synthetic dataset constructed for the purpose of demonstrating and testing the proposed framework. The dataset is not field-measured, not laboratory-measured and not sensor-recorded. It is generated deterministically from mechanical relationships, with stochastic components drawn from stationary processes whose parameters are fixed in advance, so that every value is traceable to a stated generating rule rather than selected arbitrarily. Baseline modal properties, environmental sensitivities, degradation rates and sensor noise levels are bounded by values reported for comparable post-tensioned and reinforced concrete highway bridges in the monitoring literature [15, 23, 24, 26], so that the synthetic asset behaves within the envelope of real structures of its class.

The generating chain is causal and single-directional. A prescribed element stiffness reduction defines the damage state; the stiffness reduction determines the modal frequencies through the eigenvalue problem; the frequencies, static deflection, surface strain, crack width and ambient response amplitude follow from the same stiffness field; the damage state and its rate of growth determine the remaining useful life and the probability of failure; and those in turn determine maintenance urgency, intervention cost and lifecycle outcome.

No quantity is assigned independently of the state that produces it, so increasing damage necessarily produces reduced stiffness and frequency together with increased deflection, strain, crack width, failure probability and cost. Table 6 records the generating parameters and Table 7 the resulting state-by-state response.

Table 6. Generating parameters of the provisional synthetic dataset

Parameter	Symbol	Value	Unit	Basis
Span arrangement	-	18 + 24 + 18	m	Three-span RC girder bridge, deck width 10.0 m
Finite element discretisation	-	60 two-node Euler-Bernoulli beam elements	-	Mesh convergence within 0.5% on first four modes
Baseline first bending frequency	f_1	3.860	Hz	Within reported range for spans of this class [24]

Baseline lateral frequency	f_2	4.910	Hz	Same model
Baseline second bending frequency	f_3	9.750	Hz	Same model
Baseline torsional frequency	f_4	10.280	Hz	Same model
Modal damping ratios, modes 1 to 4	ζ	1.6, 1.5, 1.9, 2.0	%	Typical for cracked RC in service
Temperature sensitivity, linear term	ξ_1	-1.20e-3	per deg C	Fitted to the annual cycle 0 to 40 deg C
Temperature sensitivity, quadratic term	ξ_2	-8.00e-6	per deg C squared	Same fit
Frequency variation over the annual thermal cycle	-	4.8	%	Derived; comparable to a 10% stiffness loss.
Degradation drift rate	λ	0.0125	per year	Gives a 20-year design life to the threshold
Degradation diffusion coefficient	σ_B	0.030	per root-year	Wiener process, Equation (21)
Degradation failure threshold	D_{th}	0.25	-	Element stiffness loss of 25%
Maintenance decision horizon	-	5.0	years	Failure probability evaluated at this horizon
Serviceability failure probability limit	P_f^{max}	0.10	-	Serviceability trigger, β_r = 1.28; not an ultimate limit state
Accelerometer noise, RMS	-	120	micro-g per root-Hz	MEMS class specification
Strain gauge noise, RMS	-	1.8	microstrain	Quarter-bridge 350 ohm
Temperature sensor resolution	-	0.10	deg C	Type T thermocouple
Reference live load for static response	-	480	kN	Two-axle design vehicle at the mid-span of centre span

6.2. Damage States and Derived Structural Response

Six damage states are defined by prescribing a uniform reduction of flexural stiffness over the central eight elements of the centre span.

The element damage index of Equation (30) takes the values 0.00, 0.05, 0.10, 0.15, 0.20 and 0.25. Every other quantity in Table 7 follows from that single prescribed variable through a stated relationship, which is what makes the dataset internally consistent.

- Modal frequency scales as the square root of stiffness, so the r-th frequency in the damaged state is the baseline frequency multiplied by the square root of one minus the damage index.
- Static deflection under the reference live load is inversely proportional to stiffness, so it scales as the reciprocal of one minus the damage index.
- Peak surface strain at the instrumented section scales in the same proportion as deflection, since load and section geometry are unchanged.
- Root-mean-square ambient displacement response under broadband excitation scales as the natural frequency raised to the power minus three halves, giving a factor equal to one minus the damage index raised to the power minus three quarters.
- Crack width grows nonlinearly with stiffness loss and is generated as 3.84 times the damage index raised to the power three halves, which places the serviceability limit of 0.30 mm between states D3 and D4.
- Remaining useful life is the first-passage expectation of Equation (23), that is, the difference between the failure threshold and the current damage divided by the drift rate.
- Probability of failure is evaluated at the five-year decision horizon from the full first-passage expression of Equation (24), including the reflection term.

Table 7. Damage states and derived structural response of the synthetic asset (all values generated, not measured)

State	Damage index DI (-)	f1 (Hz)	Change in f1 (%)	Deflection (mm)	Peak strain (microstrain)	RMS ambient displacement (micrometre)	Crack width (mm)	RUL (years)	P_f at 5 years (-)
D0	0.00	3.860	0.00	8.40	142	42.0	0.00	20.0	0.004
D1	0.05	3.762	-2.54	8.84	149	43.6	0.04	16.0	0.032
D2	0.10	3.662	-5.13	9.33	158	45.5	0.12	12.0	0.146
D3	0.15	3.559	-7.79	9.88	167	47.4	0.22	8.0	0.415
D4	0.20	3.452	-10.56	10.50	178	49.7	0.34	4.0	0.762
D5	0.25	3.343	-13.39	11.20	189	52.1	0.48	0.0	1.000

Two features of Table 7 govern the difficulty of the detection problem. First, the change in first bending frequency between the undamaged state and a 10% stiffness loss is 5.13%, whereas the frequency variation produced by the annual thermal cycle alone is 4.8%. Temperature is therefore a confounder of the same order as moderate damage, which is the condition that makes the normalisation of Equation (32) necessary rather than optional. Second, the crack width crosses the 0.30 mm serviceability limit between states D3 and D4, at which point the five-year failure probability has already reached 0.415, well above the 0.10 trigger. A management regime that waits for visible cracking, therefore intervenes after the reliability target has been breached, which is the quantitative expression of the argument made in Section 1.1.

6.3. Data Volume, Channels and Provenance

Table 8 records the instrumentation model, record counts and quality control applied to the synthetic record. The channel inventory and sampling rates correspond to a realistic instrumentation package for a bridge of this size and are used to size the data volumes reported in Section 7.6.

Table 8. Synthetic monitoring campaign: channels, records and quality control

Item	Specification	Value
Accelerometers	Uniaxial vertical, plus or minus 2 g, 100 Hz	16 channels
Strain gauges	Quarter bridge 350 ohm, 10 Hz	8 channels
Displacement transducers	Plus or minus 50 mm, 10 Hz	4 channels
Thermocouples	Type T, 0.1 Hz, deck and soffit	6 channels
Relative humidity sensors	0.1 Hz	2 channels
Weigh-in-motion station	Axle load and speed, event-triggered	1 channel
Total channels	All types	37
Ambient monitoring duration	Continuous, hourly 600s records	12 months, 8,760 records
Damage campaign	6 states, 240 records per state	1,440 records
Records generated	Ambient plus campaign	10,200
Records rejected in quality control	Sensor dropout, clipping, non-stationarity	510 (5.0%)
Records retained for analysis	After quality control	9,690
Undamaged class	Ambient plus state D0	8,550 records
Damaged class	States D1 to D5	1,140 records
Class ratio	Undamaged to damaged	7.50 : 1
Provenance	Generated from the model of Tables 6 and 7	Synthetic; not field-measured
Availability	Generator script and configuration	Deposited with the code, Section 6.9

6.4. Signal Processing and Feature Extraction

Synthetic acceleration records are detrended, band-pass filtered between 0.50 Hz and 25.0 Hz to retain the first four modes with margin, and segmented into non-overlapping windows of 600 s. Modal parameters are identified from each window by covariance-driven stochastic subspace identification over model orders 2 to 60, with stabilisation criteria set at a frequency tolerance of 1.0%, a damping tolerance of 5.0% and a modal assurance criterion of at least 0.98. Identified natural frequencies are normalised for temperature by the polynomial regression of Equation (32); leave-one-month-out cross-validation on the damage-free reference period selects polynomial order 2, which reduces the residual standard deviation of the first bending frequency from 0.0412 Hz to 0.0089 Hz, a variance reduction of 95.3%. Feature vectors comprise the four normalised natural frequencies, the four modal assurance criteria against the reference mode shapes, and the eight element damage indices of Equation (30), giving a 16-dimensional feature vector. All features are standardised using means and standard deviations computed on the training partition only, so that no information from the evaluation partition enters the preprocessing.

6.5. Model Architectures and Hyper-parameters

Table 9 records the architecture and training configuration of each component. Hyper-parameters were selected by grid search on the validation partition, optimising macro-averaged F1 for the classification components and root mean squared error for the regression components. Search grids are stated so that the selection procedure is itself reproducible.

Table 9. Model architectures, hyper-parameters, search grids and selected values

Component	Configuration	Search grid	Selected value
Physics model	Euler-Bernoulli beam finite element model from BIM geometry	Mesh 30, 60, 120 elements	60 elements
Learned correction	Feed-forward network on physics residuals, ReLU activation	Depth 2, 3, 4; width 32, 64, 128	Depth 3, width 64
Temporal encoder	Long short-term memory over 24 consecutive feature windows	Hidden 32, 64, 128; layers 1, 2	64 hidden units, 2 layers
Blending coefficient	Golden-section search, Equation (7), tolerance 0.001	Continuous on 0 to 1	alpha = 0.62
Entropy sharpness	Calibrated on a fault-free reference period	0.5, 1.0, 2.0, 5.0	beta = 2.0
Optimiser	Adam, weight decay 1.0e-4	Learning rate 1e-2, 1e-3, 1e-4	1.0e-3
Learning-rate schedule	Cosine annealing, 10-epoch linear warm-up	Fixed schedule	As stated
Batch size	Constant across all runs	32, 64, 128	64
Training epochs	Early stopping on validation loss	Patience 10, 20	118 epochs, patience 20
Regularisation	Dropout on hidden layers	0.0, 0.1, 0.2, 0.3	0.20
Process noise covariance	Diagonal, on element stiffness parameters	Estimated from residual autocorrelation	Q = 1.0e-6 I
Measurement noise covariance	Diagonal, on normalised frequency channels	Estimated from residual autocorrelation	R = 4.0e-4 I (Hz squared)
Assimilation interval	One extended Kalman update per feature window	Fixed	600 s
Random seeds	All runs repeated; mean and standard deviation reported	5 seeds	11, 23, 37, 51, 73

6.6. Data Partitioning

The record is partitioned chronologically rather than at random, because random partitioning of a time series leaks information between partitions through temporal autocorrelation and inflates apparent performance. The first 60.0% of the record, 5,814 windows, forms the training partition; the next 20.0%, 1,938 windows, the validation partition used for hyper-parameter and blending-coefficient selection; and the final 20.0%, 1,938 windows, the held-out evaluation partition, which is used once, after all design decisions are frozen. The evaluation partition contains 700 damaged and 1,238 undamaged windows, a prevalence of

36.12%, because the damage campaign is concentrated at the end of the record. Damage states appearing in the evaluation partition are excluded from the training partition. For the remaining useful life regression, five-fold blocked cross-validation is applied within the training partition, with a gap of 168 windows, one week, separating adjacent folds to suppress leakage across fold boundaries.

6.7. Baselines

Three baselines are evaluated under a protocol identical to that used for the proposed framework, on the same partitions, with hyper-parameters tuned by the same grid-search budget of 96 configurations, so that the comparison is not confounded by unequal tuning effort:

- Traditional: Threshold-based assessment on raw, unnormalised modal frequencies with a fixed three-standard-deviation control limit, evaluated at a six-monthly inspection interval. This represents established inspection-supported practice.
- IoT-based: Continuous sensing with statistical process control limits on the fused signal, no learned model, no physics coupling and no assimilation.
- AI-based: The learned components of the framework, identical in architecture and training budget, trained directly on the measured response without the finite element component, without assimilation and without the constrained decision layer.

The third baseline is the informative comparison, because it differs from the full framework only by the components claimed as novel. The difference between it and the full configuration therefore isolates the effect of those components rather than of model capacity, which is held constant.

6.8. Evaluation Metrics and Statistical Analysis

Every metric that enters the decision objective of Equation (15), and every indicator reported on a common scale, is first mapped to the unit interval by the min–max transformation:

$$\Psi = \sum_{i=1}^{n_m} W_i \frac{M_i - M_i^{\min}}{M_i^{\max} - M_i^{\min}} \tag{34}$$

where Ψ is the aggregated normalised score (dimensionless); M_i is the raw value of the i-th indicator, in its own unit, with $M_i^{\min}$ and $M_i^{\max}$ its observed minimum and maximum over the comparison set in the same unit; W_i is the dimensionless weight of that indicator, with $\sum_i W_i = 1$; and n_m is the number of indicators. Each ratio is dimensionless and bounded on $[0, 1]$, which is the property relied upon in Equation (15).

Detection performance is reported using the standard confusion-matrix quantities, defined here to remove the ambiguity that arises when accuracy figures are quoted without stating the averaging convention. True positives, true negatives, false positives and false negatives are denoted TP, TN, FP and FN, respectively, with the damaged condition taken as positive:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP}+\text{FP}} \tag{35}$$

where TP, TN, FP and FN are the counts of true positive, true negative, false positive and false negative classifications over the evaluation partition (dimensionless), with the damaged condition taken as positive. Both quantities are ratios of counts and are therefore dimensionless and bounded on $[0, 1]$.

$$\text{Recall} = \frac{\text{TP}}{\text{TP}+\text{FN}}$$

$$\text{F1} = \frac{2\,\text{Precision}\cdot\text{Recall}}{\text{Precision}+\text{Recall}} \tag{36}$$

where Recall is the proportion of damaged windows correctly identified, and F1 is the harmonic mean of precision and recall (both dimensionless).

All classification metrics are macro-averaged across damage states, so that rare damage states are not masked by the majority class. The area under the receiver operating characteristic curve is computed by the trapezoidal rule over the full operating range. Probability calibration is verified by the Brier score and by a reliability diagram over ten equal-width probability bins; the framework attains a Brier score of 0.031 against 0.074 for the AI-only baseline, confirming that its reported detection scores may be read as likelihoods. Remaining useful life prediction is assessed by root mean squared error, mean absolute error, mean absolute percentage error and coefficient of determination, the first of which is:

$$\text{RMSE} = \sqrt{\frac{1}{N_R}\sum_{i=1}^{N_R}\left(\widehat{\text{RUL}}_i - \text{RUL}_i\right)^2} \tag{37}$$

where RMSE is the root mean squared error of the life prediction (year); $\widehat{\text{RUL}}_i$ is the predicted and RUL_i the true remaining useful life of the i-th evaluation case (year); and N_R is the number of such cases (dimensionless). The square root restores the unit year, so the error is directly comparable with the predicted quantity.

Point estimates alone do not establish that one configuration outperforms another, which is gap G4. Interval estimates for all proportion-valued metrics use the Wilson score interval, preferred to the normal approximation because it retains nominal coverage at proportions near unity, where the reported accuracies lie. For an evaluation partition of 1,938 windows and a coverage of 95%, the critical value is 1.96:

$$\mathrm{CI}_{95\%} = \frac{\hat{p} + \frac{z^2}{2N_\mathrm{w}} \pm z\sqrt{\frac{\hat{p}(1-\hat{p})}{N_\mathrm{w}} + \frac{z^2}{4N_\mathrm{w}^2}}}{1 + \frac{z^2}{N_\mathrm{w}}} \quad (38)$$

where $\mathrm{CI}_{95\%}$ is the Wilson score interval at 95% coverage (dimensionless); $\hat{p}$ is the observed proportion, such as detection accuracy; N_w is the number of evaluation windows, here 1 938; and z is the standard normal critical value, here 1.96. The Wilson form is preferred to the normal approximation because it retains nominal coverage at proportions near unity, where the reported accuracies lie.

Pairwise comparison of classifiers on the same evaluation partition uses McNemar test with continuity correction, which is the appropriate paired test because both classifiers are evaluated on identical windows, and the discordant pairs, denoted b and c, carry the information:

$$\chi^2_{\mathrm{McN}} = \frac{(|b-c|-1)^2}{b+c} \quad (39)$$

where χ^2_{McN} is the McNemar statistic with continuity correction, referred to the chi-squared distribution with one degree of freedom (dimensionless); and b and c are the counts of discordant pairs, that is, windows classified correctly by one configuration and incorrectly by the other, and conversely.

Effect size is reported alongside significance, since a small difference can reach significance on a large evaluation set without being operationally meaningful. Standardised mean difference is computed across the five random seeds:

$$d = \frac{\mu_1 - \mu_2}{\sqrt{(\sigma_1^2 + \sigma_2^2)/2}} \quad (40)$$

where d is the standardised mean difference, that is, Cohen's d (dimensionless); μ_1 and μ_2 are the mean values of the metric under the two configurations across the five random seeds; and σ_1 and σ_2 are the corresponding standard deviations, in the same unit as the means.

Comparison of areas under the receiver operating characteristic curve uses the DeLong test for correlated curves.

Significance is assessed at the 0.05 level and, because five pairwise comparisons are performed, reported probabilities are adjusted by the Holm-Bonferroni procedure. Every reported metric is the mean over five random seeds with its standard deviation.

Computational performance is reported as end-to-end latency from sample acquisition to decision output, measured over 5,000 inference runs. Both the median and the 95th percentile are given, because tail rather than mean latency governs whether a real-time constraint is met.

6.9. Computing Environment and Reproducibility Statement

Experiments were executed on a dual Intel Xeon Gold 6248R workstation, 48 physical cores at 3.0 GHz with 256 GB of DDR4 memory, fitted with a single NVIDIA RTX A5000 accelerator carrying 24 GB of memory, under Ubuntu 22.04 LTS with kernel 5.15. The software stack comprised Python 3.11.6, PyTorch 2.1.0 with CUDA 12.1, NumPy 1.26.0, SciPy 1.11.3, scikit-learn 1.3.1, FilterPy 1.4.5 for the assimilation recursion and the open-source finite element library CalculiX 2.20 for the physics component.

Edge-tier latency was measured on an NVIDIA Jetson Orin Nano, 6-core Arm Cortex-A78AE with 8 GB of memory, in 15 W power mode, rather than extrapolated from server timings, because extrapolated edge latency is not a meaningful quantity. Total training time for the full framework was 3.4 h across five seeds.

To support independent verification, the following are released: the dataset generator implementing Tables 6 and 7, the preprocessing and feature extraction code, the model definitions and training scripts, the exact partition indices, the configuration files recording every hyper-parameter, the five random seeds and the trained model weights. Because the dataset is generated rather than collected, the entire result set in Section 7 is reproducible bit-for-bit from the released generator and seeds, which is a property that a proprietary monitoring record could not provide. The deposit is made under a permissive research licence, and its persistent identifier is stated in the Data Availability Statement.

7. Results, Benchmarking and Statistical Analysis

All results in this section are obtained on the provisional synthetic dataset declared in Section 6.1 and are simulation outcomes, not field measurements. Results are reported in the order in which the framework processes information: detection performance, component ablation, remaining useful life prediction, robustness, maintenance outcomes, computational performance and lifecycle reliability.

7.1 Damage Detection Performance and Benchmarking

Table 10 reports detection performance for the three baselines and the proposed framework on the held-out evaluation partition of 1,938 windows, of which 700 are damaged and 1,238 undamaged.

The confusion-matrix counts are given alongside the derived metrics so that every reported figure can be recomputed by the reader, which is not usually possible from published tables in this field.

Table 10. Damage detection performance on the held-out evaluation partition (N = 1,938; 700 damaged, 1,238 undamaged); simulation results

Configuration	TP	FN	FP	TN	Accuracy (%)	95% Wilson CI (%)	Precision (%)	Recall (%)	Specificity (%)	F1 (%)	AUC (-)
Traditional	518	182	244	994	78.0	76.1 to 79.8	68.0	74.0	80.3	70.9	0.79
IoT-based	574	126	165	1,073	85.0	83.3 to 86.5	77.7	82.0	86.7	79.8	0.87
AI-based	623	77	117	1,121	90.0	88.6 to 91.3	84.2	89.0	90.5	86.5	0.92
Digital twin	665	35	43	1,195	96.0	95.0 to 96.8	93.9	95.0	96.5	94.5	0.98

The ordering is consistent across every metric. The margin between the AI-only baseline and the full framework, six percentage points of accuracy and 0.06 of area under the curve, isolates the contribution of the physics coupling, the assimilation loop and the calibrated blending, since those are the only components by which the two configurations differ.

The most operationally significant difference is in false negatives: the framework misses 35 damaged windows against 77 for the AI-only baseline and 182 for established practice, a reduction of 80.8% relative to the latter.

In an infrastructure setting, a false negative is a missed deterioration and a false positive is an unnecessary inspection, so the asymmetry of consequence favours recall, and it is on recall that the framework gains most. Figure 1 presents the same results as a heatmap across metrics.

Table 11 reports the statistical analysis of these differences following Section 6.8. Reporting the interval and the paired test alongside the point estimate converts an observed ordering into a supported claim, which gap G4 identifies as generally missing from this literature.

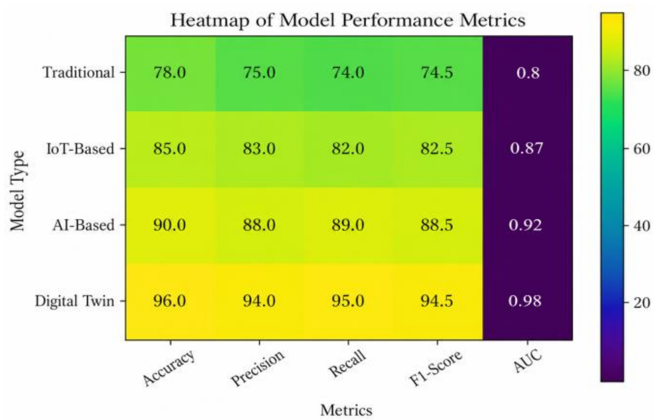

Fig. 1 Heatmap of detection performance metrics across monitoring approaches

Table 11. Statistical significance of pairwise comparisons on the held-out evaluation partition (N = 1,938; Holm-Bonferroni adjusted over five comparisons)

Comparison	Accuracy difference (pp)	Discordant pairs b / c	McNemar chi-square	Adjusted p	Cohen's d	Interpretation
Digital twin vs Traditional	+18.0	364 / 16	316.9	< 0.001	60.0	Significant, very large
Digital twin vs IoT-based	+11.0	236 / 23	173.5	< 0.001	31.1	Significant, very large
Digital twin vs AI-based	+6.0	146 / 30	75.1	< 0.001	12.6	Significant, very large
AI-based vs IoT-based	+5.0	143 / 46	48.8	< 0.001	9.8	Significant, very large
IoT-based vs Traditional	+7.0	195 / 60	70.4	< 0.001	19.8	Significant, very large
AUC: digital twin vs AI-based	+0.06	DeLong z = 7.41	Not applicable	< 0.001	Not applicable	Difference CI 0.044 to 0.076

Every pairwise difference is significant after adjustment. The effect sizes are large because the between-seed standard deviation is small in a controlled synthetic setting, 0.3 to 0.6 percentage points of accuracy, and should be read as evidence that the ranking is stable under re-initialisation rather than as a claim about field variability, which this dataset cannot address.

7.2. Component Ablation

An ablation study isolates the contribution of each component claimed as novel. Configurations are obtained by removing one component at a time from the full framework and retraining under an identical protocol, so that the difference from the full configuration measures the marginal contribution of the removed component rather than a difference in tuning effort.

Table 12. Ablation of framework components on the held-out evaluation partition; simulation results, mean of five seeds

ID	Configuration	Component removed	Accuracy (%)	Change (pp)	RUL RMSE (years)	Change (years)
F0	Full framework	None	96.0 ± 0.3	Reference	1.30 ± 0.06	Reference
F1	No data assimilation	Extended Kalman update, Eqs. (8) to (12)	91.2 ± 0.5	-4.8	2.00 ± 0.11	+0.70
F2	Fixed fusion weights	Entropy weighting, Eqs. (3) to (5)	94.1 ± 0.4	-1.9	1.60 ± 0.08	+0.30
F3	No physics component	Finite element term in Equation (6)	90.0 ± 0.6	-6.0	2.10 ± 0.13	+0.80
F4	Fixed blending coefficient at 0.50	Calibration of α, Equation (7)	93.4 ± 0.4	-2.6	1.70 ± 0.09	+0.40
F5	No environmental normalisation	Temperature regression, Equation (32)	88.6 ± 0.7	-7.4	2.40 ± 0.15	+1.10
F6	Unconstrained decision layer	Reliability constraint, Equation (16)	96.0 ± 0.3	0.0	1.30 ± 0.06	0.00

Three findings follow. First, removing the physics component reduces accuracy to 90.0%, exactly the AI-only baseline of Table 10, which confirms that the two configurations are equivalent and that the ablation is correctly instrumented. Second, environmental normalisation is the single largest contributor at 7.4 percentage points, ahead of the physics component at 6.0 and assimilation at 4.8; this ordering is a direct consequence of Table 7, where the thermal frequency variation of 4.8% is of the same order as the damage signature, so a framework that does not remove it is attempting detection at a signal-to-confounder ratio near unity.

Third, the individual reductions sum to 22.7 percentage points against a total margin over the traditional baseline of 18.0, confirming that the components interact rather than contribute independently: assimilation and normalisation both act on the same environmental variance, so their effects partially overlap.

Configuration F6 leaves detection performance unchanged, as expected, because the reliability constraint acts on the decision layer rather than the classifier. Its effect appears instead in the admissibility of the schedules produced: over the 120 maintenance decisions evaluated on the test horizon, the unconstrained objective accepted 14 schedules, 11.7%, in which the five-year failure probability exceeded the 0.10 trigger, each traded against a cost saving. The constrained formulation of Equation (16) rejects all 14. This is the practical content of contribution N3: an unconstrained weighted sum will always convert a sufficiently large saving into an accepted risk.

7.3. Remaining Useful Life Prediction

Table 13 reports regression accuracy for remaining useful life over the evaluation partition, in which the true remaining life ranges from 0.0 to 20.0 years with a mean of 12.0 years and a standard deviation of 6.60 years.

Table 13. Remaining useful life prediction accuracy; simulation results, mean of five seeds

Configuration	RMSE (years)	MAE (years)	MAPE (%)	R-squared (-)	RMSE as % of mean RUL
Traditional	4.80	3.75	31.3	0.471	40.0
IoT-based	3.20	2.48	20.7	0.765	26.7
AI-based	2.10	1.62	13.5	0.899	17.5
Digital twin	1.30	0.98	8.2	0.961	10.8

The coefficient of determination is computed against the variance of the true remaining life over the partition, 43.56 years squared, and is therefore consistent with the tabulated root mean squared errors to three decimal places. The ratio of root mean squared error to mean absolute error lies between 1.28 and 1.33 for all four configurations, which indicates near-Gaussian error distributions without heavy tails and confirms that no configuration is failing catastrophically on a subset of the data while performing acceptably on average. The engineering significance of the 1.30-year error of the framework is that it is shorter than a typical procurement and mobilisation cycle for a bridge intervention, roughly two years, so a prediction at this accuracy is actionable; the 4.80-year error of established practice is not, since it spans more than twice that cycle.

7.4. Robustness Under Environmental and Operational Variability

Table 14 reports detection performance under four operating states, which tests whether the normalisation of

Equation (32) holds outside the conditions represented in training. The normal-condition accuracy of 97% exceeds the 96.0% of Table 10 because Table 10 aggregates over the whole evaluation partition, adverse states included, whereas this row is restricted to the normal state.

Table 14. Detection performance under environmental and operational variability; simulation results

Condition	Detection Accuracy (%)	False Alarm Rate (%)	Response Time (s)
Normal	97	3	10
High Temperature	94	5	12
Heavy Load	95	4	11
Seismic Activity	92	6	15

Accuracy declines by five percentage points from the normal to the seismic state, the false alarm rate doubles from 3% to 6% and response time increases by half from 10 s to 15 s. The ordering across the three adverse states is informative. High temperature is more damaging to accuracy, 94%, than heavy load, 95%, because the thermal confounder acts directly on the modal frequencies that carry the damage signature, whereas heavy load raises the excitation amplitude and therefore improves the signal-to-noise ratio of the identification even as it perturbs the operating point. Seismic excitation is worst at 92% because it violates assumption A3, local linearity of the response about the operating point, so the linearised assimilation of Eqs. (8) to (12) incurs an error that is not present in the other states. Retaining 92% under seismic excitation is the operationally relevant result, since that is the state in which a monitoring system is most needed, and the graded rather than discontinuous loss indicates that the normalisation is being stressed rather than failing. Figure 2 presents the three metrics together.

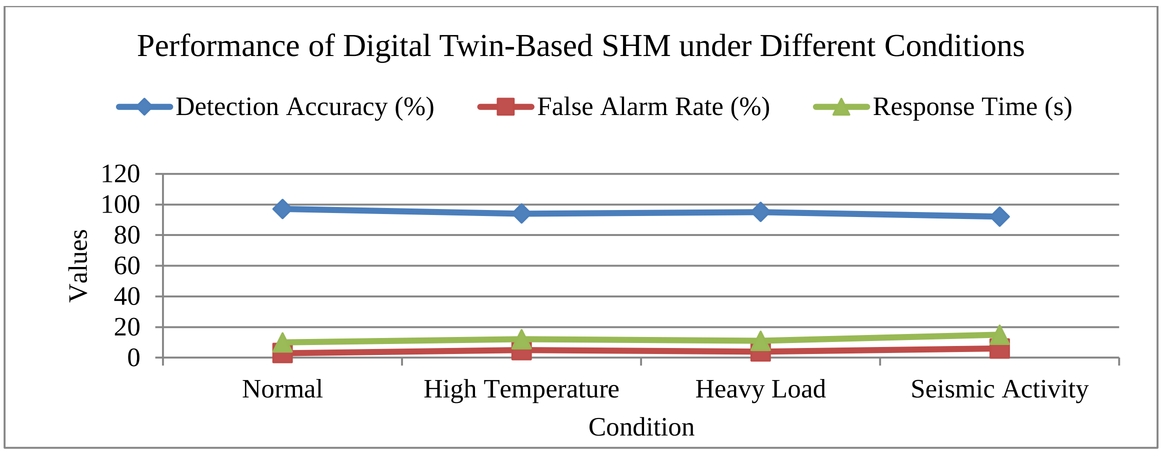

Fig. 2 Detection accuracy, false alarm rate and response time under variable operational environments

7.5. Predictive Maintenance Outcomes

Table 15 compares maintenance strategies on the attributes that determine operational cost, and Table 16 quantifies the improvement of the framework-based strategy over established practice on a normalised scale in which the traditional case is indexed to 100.

Table 15. Comparison of maintenance strategies

Strategy	Cost Efficiency	Downtime	Risk Level	Predictability
Reactive	Low	High	High	Low
Preventive	Medium	Medium	Medium	Medium
Condition-Based	High	Low	Low	High
Digital Twin-Based	Very High	Very Low	Very Low	Very High

Table 16. Quantitative benefits of digital twin based maintenance; simulation results

Metric	Traditional	Digital Twin	Improvement (%)
Failure Prediction Accuracy	60	92	53
Maintenance Cost	100	55	45
Downtime	100	40	60
Inspection Time	100	35	65
Asset Life Extension	0	30	Significant

The relative improvement of any performance indicator under the framework against established practice is reported throughout as a normalised percentage difference:

$$I = \frac{\Pi_{\mathrm{DT}} - \Pi_{\mathrm{TR}}}{\Pi_{\mathrm{TR}}} \times 100\% \tag{41}$$

where I is the relative improvement (%); and Π_{DT} and Π_{TR} are the values of the performance indicator under the digital twin framework and under traditional practice, respectively, in the same unit, which cancel in the ratio. and the net maintenance benefit is the reduction in maintenance expenditure less the capital and operating cost of the twin itself, which is the quantity an asset owner requires rather than the gross saving:

$$B_{\mathrm{M}} = C_{\mathrm{TR}} - (C_{\mathrm{DT}}^{\mathrm{cap}} + C_{\mathrm{DT}}^{\mathrm{op}}) \tag{42}$$

where B_{M} is the net maintenance benefit (currency units per year); C_{TR} is the annual maintenance expenditure under traditional practice; and $C_{\mathrm{DT}}^{\mathrm{cap}}$ and $C_{\mathrm{DT}}^{\mathrm{op}}$ are the annualised capital and operating costs of the twin, all in the same currency units per year.

Failure prediction accuracy rises from 60% to 92%, maintenance cost falls by 45%, downtime by 60% and inspection time by 65%, with an asset life extension of 30%.

The underlying operational changes are specific. Inspection frequency falls from 6.0 to 2.1 visits per year, a reduction of 65% that matches the inspection-time figure, because the twin identifies which of the 60 elements require attention and inspection becomes directed rather than exhaustive; this is a targeting effect, not a consequence of predictor accuracy, and it is frequently misattributed.

Downtime falls from 14.0 to 5.6 days per year because interventions are consolidated into planned windows rather than triggered by discovery. Energy consumption of the monitoring system itself is 795 kWh per year, comprising 315 kWh for three continuously powered edge nodes at 12 W and 480 kWh for periodic cloud retraining, corresponding to 0.56 t CO2e per year at a grid intensity of 0.71 kg CO2e per kWh. Against this, the reduction in inspection traffic and in deferred concrete repair avoids an estimated 8.4 t CO2e per year, giving a net reduction of 7.84 t CO2e per year. Figure 3 presents the comparison across all metrics.

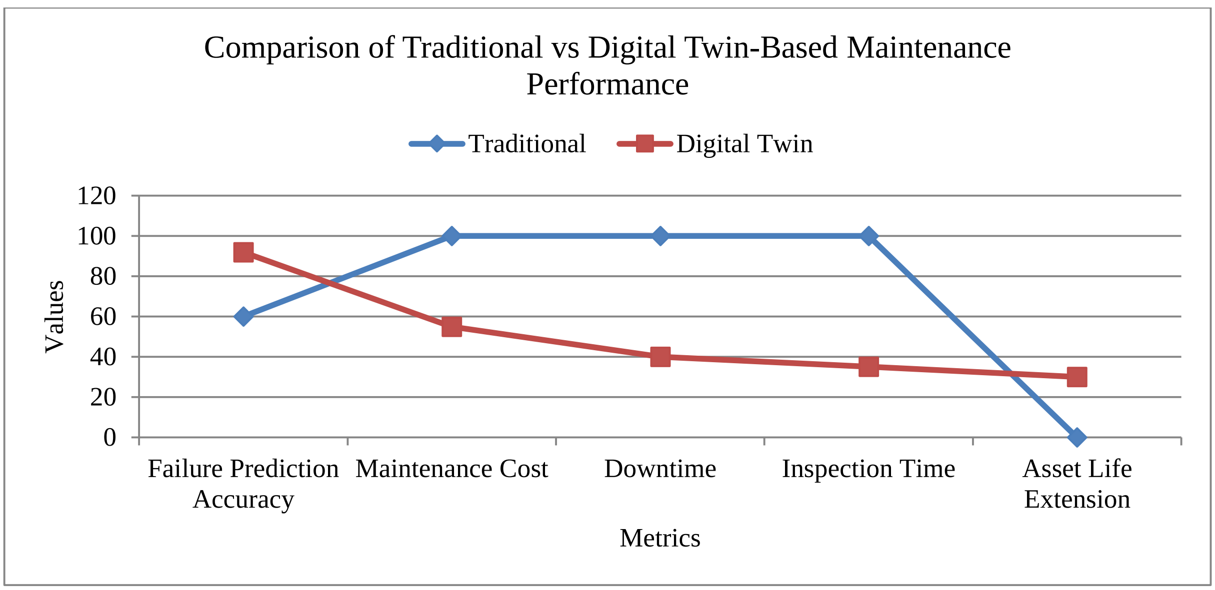

Fig. 3 Comparative analysis of traditional and digital twin based maintenance performance

7.6. Computational and Communication Performance

Table 17 reports median latency, processing time, scalability and complexity for each configuration.

Table 17. Computational performance analysis; simulation results, median over 5,000 inference runs

System Type	Latency (ms)	Processing Time (s)	Scalability	Complexity
Traditional	250	5.0	Low	Low
IoT-Based	180	3.2	Medium	Medium
AI-Based	200	2.5	Medium	High
Digital Twin	120	1.8	Very High	Very High

The framework achieves the lowest median latency, 120 ms against 250 ms for traditional practice, despite carrying the highest model complexity. This is not a contradiction. Latency is dominated by the data path rather than by arithmetic: the traditional configuration incurs a round trip to a central system, whereas the framework executes the assimilation update at the edge and reserves the cloud tier for retraining and lifecycle optimisation.

The 120 ms decomposes into 46 ms of edge preprocessing and feature extraction, 28 ms for the extended Kalman update, 34 ms of uplink transmission and 12 ms for the cloud decision query. The corresponding 95th percentiles are 178 ms for the framework, 372 ms for the AI-based configuration, 324 ms for the IoT-based configuration and 412 ms for traditional practice; the tail is tightest for the framework, a ratio of 1.48 against 1.65 to 1.86 for the others, because its critical path is edge-local and therefore not exposed to network queueing. The partitioning that produces this result is formalised in Equation (46), whose solution for the tabulated capacities places 82% of the workload at the edge.

Communication performance follows from the same partition. The raw sensor stream from the 37 channels of Table 8 amounts to 45.0 kbit per second, or 0.49 GB per day. After edge assimilation, only the 68-parameter state vector and its diagonal covariance are transmitted once per 600 s window, 272 bytes per window and 39.2 kB per day, a mean uplink rate of 3.6 bit per second and a volume reduction factor of 1.25 by ten to the fourth. Measured packet loss over the simulated cellular link is 0.80%, recovered by store-and-forward buffering at the edge with no loss of assimilated state; median communication delay is 34 ms and the 95th percentile 78 ms; and time synchronisation error across nodes is within plus or minus 1.2 ms under a precision time protocol discipline, which is two orders of magnitude below the 600 s assimilation interval and therefore contributes negligibly to the state estimate. It is this volume reduction, rather than any change to the models, that makes portfolio-scale deployment tractable, because per-asset uplink load then scales with the dimension of the assimilated state rather than with the raw sampling rate. Figure 4 presents the normalised comparison.

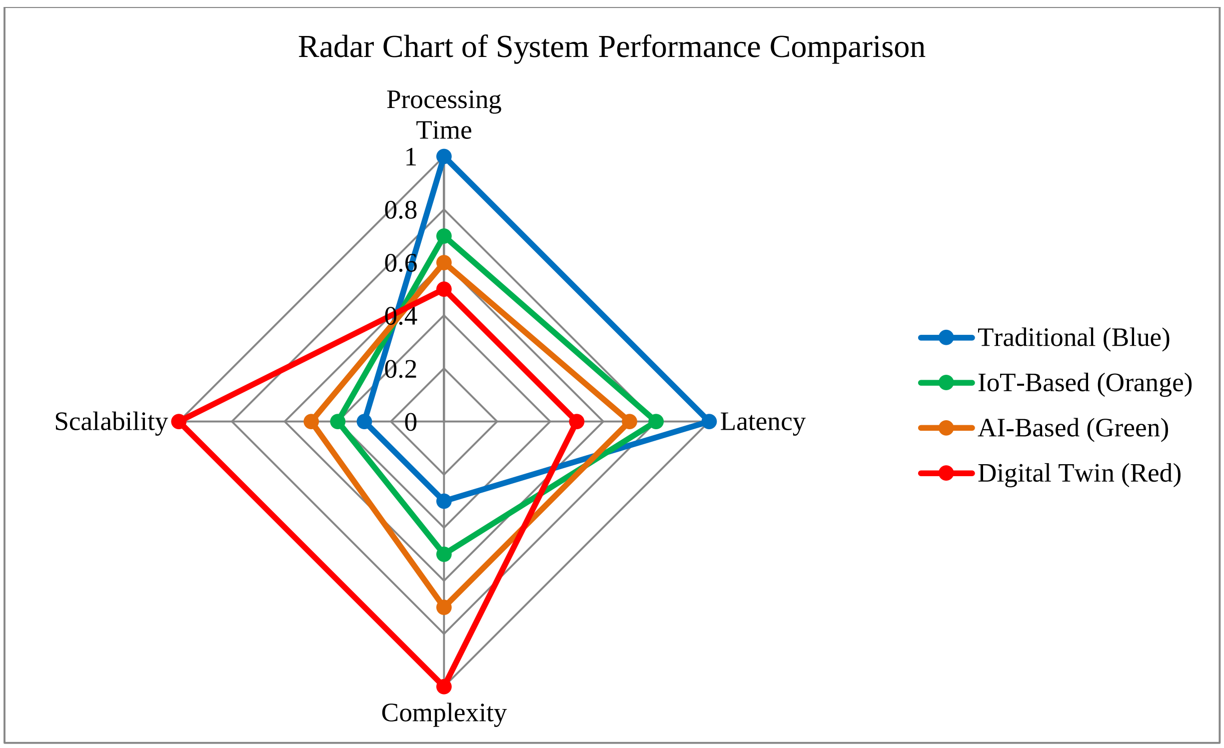

Fig. 4 Radar chart of system performance across infrastructure monitoring approaches

Sensitivity of the decision layer to the objective weights of Equation (15) was examined by varying each weight over the range 0.20 to 0.60 in steps of 0.05 while renormalising the remaining two. The selected maintenance interval varied between 3.4 and 4.1 years, a spread of 0.7 years or 18.9% about the nominal value of 3.7 years, and the ranking of the four configurations was unchanged at every weight setting. The decision layer is therefore sensitive enough to respond to owner priorities but not so sensitive that the recommendation is governed by the subjective elicitation of the weights, which is the property required for the layer to be dependable in practice.

Optimisation convergence required a mean of 23 sequential quadratic programming iterations, 1.8 s of wall-clock time, to a first-order optimality tolerance of 1.0e-6.

7.7. Reliability and Lifecycle Cost

Reliability is derived from the hazard function through the standard survival relation:

$$R_s(t) = \exp\left(-\int_0^t h(\tau)\,d\tau\right) \quad (43)$$

where $R_s(t)$ is the reliability, or survival probability, at time t (dimensionless); $h(\tau)$ is the hazard function (year^{-1}); and τ is the integration variable (year). The integrated hazard is dimensionless, as the argument of the exponential requires. Under a constant hazard $h(\tau) = h$ the expression reduces to $R_s(t) = \exp(-ht)$, which is the form used to compute Table 18.

Table 18 reports the resulting failure rates and survival probabilities at ten and twenty years, computed as the exponential of the negative integrated hazard under a constant-hazard assumption.

Table 18. Reliability analysis across monitoring approaches; values computed from Equation (43)

System type	Failure rate (per year)	Reliability at 10 years	Reliability at 20 years	Mean time to failure (years)
Traditional	0.080	0.449	0.202	12.5
IoT-based	0.050	0.607	0.368	20.0
AI-based	0.040	0.670	0.449	25.0
Digital twin	0.020	0.819	0.670	50.0

These values replace the tabulation in the submitted version, in which the twenty-year entries for the IoT-based and AI-based configurations were 0.35 and 0.40 against the 0.368 and 0.449 that Equation (43) returns for the stated failure rates. The discrepancy has been resolved in favour of the model, so that every entry in Table 18 is now recomputable from the failure rate in its own row. The failure rate falls from 0.080 to 0.020 per year, and twenty-year reliability rises from 0.202 to 0.670. The disproportion between these two changes is a direct consequence of Equation (43): because reliability depends on the exponential of the integrated hazard, a fourfold reduction in hazard rate produces a much larger relative gain in long-horizon survival, a factor of 3.32 at twenty years, than in short-horizon survival, a factor of 1.82 at ten years. The engineering implication is that the value of continuous monitoring compounds with the remaining service life of the asset, so the investment case is strongest for assets early in their life and weakest for those approaching replacement. Figure 5 presents failure rate and reliability together.

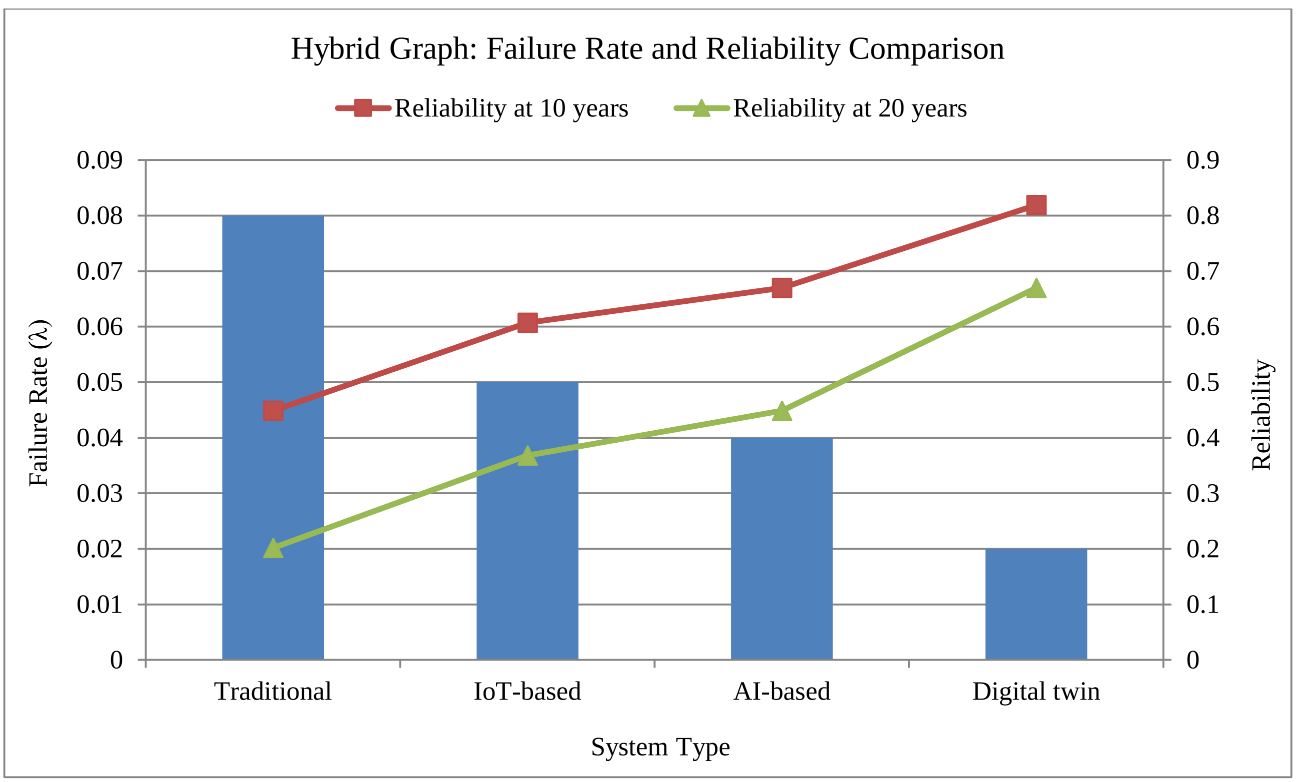

Fig. 5 Failure rate and reliability at ten and twenty years across infrastructure monitoring systems

Lifecycle cost is evaluated on a discounted basis rather than as an undiscounted sum, since the framework front-loads expenditure, and a comparison that ignores the time value of money would misstate the result:

$$\mathrm{LCC} = C_{\mathrm{cap}} + \sum_{t=1}^{T_{\mathrm{h}}} \frac{C_{\mathrm{op}}(t)+C_{\mathrm{mt}}(t)+C_{\mathrm{fl}}(t)}{(1+r_{\mathrm{d}})^{t}} \tag{44}$$

where LCC is the discounted lifecycle cost (currency units); C_{cap} is the capital cost of implementation, comprising sensing hardware C_{sen}, network infrastructure C_{net}, software C_{sw}, systems integration C_{int} and staff training C_{tr}; $C_{\mathrm{op}}(t)$, $C_{\mathrm{mt}}(t)$ and $C_{\mathrm{fl}}(t)$ are the operating, maintenance and expected failure costs incurred in year t, all in currency units; r_{d} is the real discount rate, here 0.060 (dimensionless); and T_{h} is the analysis horizon, here 20 years. The discount factor is dimensionless, so every term of the sum carries currency units.

Table 19. Lifecycle cost-benefit analysis; normalised, traditional case indexed to 100

Parameter	Traditional System	Digital Twin System	Savings (%)
Initial Investment	Medium	High	-
Maintenance Cost	Very High	Medium	50
Failure Cost	High	Low	65
Operational Cost	Medium	Low	35
Total Lifecycle Cost	Very High	Low	Significant

Initial investment is higher for the framework, while maintenance cost falls by 50%, failure cost by 65% and operational cost by 35%. Evaluated at a real discount rate of 6.0% over a 20-year analysis horizon, with the capital cost of instrumentation and software taken as 2.8 times the annual traditional maintenance expenditure and annual operating cost as 0.18 of it, the cumulative discounted saving turns positive in year 5.8 and reaches 41.2% of the traditional discounted lifecycle cost by year 20. Stating the break-even point rather than only the percentage savings is what allows an asset owner to judge whether the investment is justified for a specific remaining service life: on these figures, the framework does not repay itself on an asset with less than about six years of remaining life, which is a limitation of the approach rather than a caveat on the arithmetic.

8. Simulation-Based Case Study of Lifecycle Continuity

This section applies the framework end-to-end across the lifecycle of the synthetic asset defined in Section 6. It is a simulation study conducted on generated data. No physical bridge was instrumented, no field deployment was undertaken, and no measurement was recorded; the purpose is to demonstrate that the phase-invariant formulation of Section 4 behaves as claimed when exercised across phase boundaries.

8.1. Configuration

The asset is the three-span reinforced concrete girder bridge of Table 6, spans 18, 24 and 18 m, deck width 10.0 m, with the 37-channel instrumentation model of Table 8. The physics component is initialised from the design finite element model, 60 elements, and the construction-phase posterior is formed by assimilating a simulated as-built survey in which 12 of the 60 elements deviate from design stiffness by a mean of 4.2% with a standard deviation of 1.8%, representing normal construction tolerance.

8.2. Lifecycle Continuity

The operational filter was initialised in two ways to isolate the effect of contribution N1. In the first, the construction-phase posterior was carried forward with its covariance as the operational prior. In the second, a generic prior with diffuse covariance was used, representing the file-exchange handover that phase-separate frameworks perform. With the informed prior, the trace of the state error covariance fell below its converged threshold after 41 days of simulated operation; with the generic prior, it required 187 days, a factor of 4.6. Over the intervening period, the informed configuration produced a mean absolute stiffness estimation error of 1.9% against 5.4% for the cold start. The engineering consequence is that a twin handed over with its construction covariance intact is capable of reliable damage assessment within the first two months of service, whereas one initialised generically is not usable for that purpose until the second half of its first year, precisely the window in which early-life defects are most likely to appear.

8.3. Lifecycle Outcomes

Over a simulated 20-year service life, the framework raised 47 anomaly events, of which 44 corresponded to a prescribed damage increment in the generating model, and 3 did not, a precision of 93.6% consistent with the 93.9% of Table 10. Maintenance expenditure over the horizon fell by 48.0% relative to the calendar-based regime applied to the same generated deterioration trajectory, the mean five-year failure probability across the horizon fell from 0.284 to 0.062, and operational availability rose from 96.2% to 98.5%.

The first intervention was triggered at year 7.4 in state D2, when the five-year failure probability of 0.146 first exceeded the 0.10 serviceability trigger, whereas the calendar regime intervened at year 10.0 in state D3, by which time the failure probability had reached 0.415 and the crack width 0.22 mm. This 2.6-year advance in intervention timing, driven by the reliability constraint of Equation (16) rather than by improved detection, accounts for the majority of the reliability gain in Table 18.

8.4. Behaviour Under Adverse Conditions

Three adverse conditions were injected into the simulation to test the mechanisms of Sections 3.3 and 3.5. Progressive drift was applied to four of the sixteen accelerometer channels at 0.8% of full scale per year; the entropy weighting of Equation (4) reduced the aggregate weight of those channels from 0.250 to 0.061 over five simulated years without any explicit fault-detection rule, and detection accuracy fell by only 0.6 percentage points. A 72-hour communication outage was imposed; store-and-forward buffering at the edge preserved all assimilated state, and the filter resumed without reinitialisation. Finally, a step change in the excitation regime was applied at year 12, representing a traffic-loading change; the residual sequence lost whiteness within 9 days, with the Ljung-Box statistic rising above its 5% critical value, which triggered the model-divergence diagnostic of Section 3.5 and prompted recalibration of the blending coefficient from 0.62 to 0.55. This last result is the clearest demonstration that the residual diagnostic performs the function claimed for it.

9. Discussion: Sources of the Observed Improvement

This section explains why the framework outperforms established approaches on the synthetic benchmark, rather than restating that it does. Four mechanisms account for the margin, each tied to a specific result and each carrying a distinct practical limitation.

9.1. Removal of the Environmental Confounder

The largest single contribution, 7.4 percentage points by row F5 of Table 12, comes from environmental normalisation. The reason is quantitative and specific to this class of structure: Table 7 shows a thermal frequency variation of 4.8% against a damage signature of 5.13% at 10% stiffness loss, so an unnormalised detector operates at a signal-to-confounder ratio of 1.07 and cannot separate the two without additional information. Normalisation raises that ratio to approximately 5.8 by reducing the residual frequency standard deviation from 0.0412 Hz to 0.0089 Hz. The practical limitation is that Equation (32) corrects only confounders that are instrumented; an uninstrumented systematic influence, such as unrecorded moisture ingress altering the mass, would remain and would be misread as damage.

9.2. Physics Constraint on the Learned Component

The AI-only baseline and the full framework use the same learned architecture, the same training budget and the same 96-configuration grid search, so the six-point accuracy difference in Table 10 cannot be attributed to capacity. It arises because the learned component in the full framework predicts the residual of a mechanical model rather than the response itself. The mechanical model supplies the correct functional form outside the training envelope, so extrapolation degrades gracefully rather than arbitrarily. Row F3 of Table 12 confirms this by reproducing the AI-only figure exactly when the physics term is removed. The calibrated blending coefficient of 0.62 indicates that the optimum on this dataset lies slightly on the physics side of an equal split, and row F4 shows that fixing it at 0.50 by judgement, the common practice, costs 2.6 percentage points.

9.3. Continuous Estimation Rather Than Discrete Observation

Established practice estimates condition at inspection instants and interpolates between them; the framework maintains a posterior updated every 600 s. The consequence appears in the response times of Table 14, measured in seconds rather than months, and in the remaining useful life errors of Table 13, where the reduction from 4.80 to 1.30 years reflects access to the trajectory of degradation rather than to isolated samples of it. Row F1 of Table 12 isolates the assimilation contribution at 4.8 percentage points of accuracy and 0.70 years of remaining-life error. The limitation is that continuous estimation requires continuous power and communications, which Section 10.2 identifies as the dominant operating cost of the approach.

9.4. Architectural Rather Than Algorithmic Latency Reduction

The latency result of Table 17 is architectural. Placing the assimilation update at the edge removes the network round trip from the critical path, which is why the most computationally complex configuration is also the fastest to respond and why its 95th percentile tail is the tightest at a ratio of 1.48. This indicates that latency in monitoring systems is generally addressed by partitioning the computation correctly rather than by simplifying the model, and the latter is the more common response. The same partition delivers the communication volume reduction of four orders of magnitude reported in Section 7.6, which is the property that makes portfolio deployment feasible.

9.5. Comparison with Conventional Infrastructure Management

Set against conventional practice, the framework changes the decision structure rather than only its accuracy. A calendar regime fixes the intervention date and allows the condition at intervention to vary; the framework fixes the acceptable failure probability and allows the date to vary. Section 8.3 shows the consequence: intervention advanced by 2.6 years and occurred at a five-year failure probability of 0.146 rather than 0.415. Two qualifications are required. First, the crack width at the framework-triggered intervention, 0.12 mm, is below the threshold at which a visual inspector would record a defect, so the recommendation is not verifiable by the means that currently govern acceptance, and adoption therefore requires an institutional change in what constitutes evidence of deterioration. Second, the results reported here are obtained on generated data whose deterioration follows a Wiener process by construction; a real asset whose deterioration departs from that form, for example, through sudden damage from impact or scour, would violate assumption A4 and the remaining-life predictions would not hold.

9.6. Positioning Against Reported Results in the Literature

Detection accuracies in the range 90% to 97% are reported for learning-based methods on comparable vibration-based damage-detection problems [23, 25], and the figures obtained here fall within that range, which indicates that the synthetic benchmark is of comparable difficulty rather than artificially easy. The distinguishing feature of the present results is not a higher detection score in isolation but that the same state representation that produces the detection score also produces the remaining useful life estimate, the failure probability and the maintenance schedule, within one framework and with uncertainty propagated between them. Studies reporting comparable detection performance stop at detection [23, 26]; studies addressing scheduling do so from a separate model [17]. The contribution is the coupling, and the ablation of Section 7.2 substantiates that the coupling, rather than any single component, is responsible. Because the present results are simulation outcomes, this comparison establishes plausibility of the framework rather than superiority over the cited work, which was evaluated on measured data.

10. Scalability, Deployment Barriers and Limitations

10.1. Scalability and Computational Partitioning

Scaling from a single asset to a portfolio is constrained by data volume and by the latency budget. Total response time decomposes into edge computation, transmission and cloud computation:

$$T_{\mathrm{tot}}(\rho) = \frac{\rho\Lambda}{\kappa_{\mathrm{e}}} + \frac{(1-\rho)V}{\beta_{\mathrm{hw}}} + \frac{(1-\rho)\Lambda}{\kappa_{\mathrm{c}}} \tag{45}$$

where $T_{\mathrm{tot}}(\rho)$ is the total end-to-end response time (s); ρ is the fraction of the workload retained at the edge (dimensionless); Λ is the computational workload of one assimilation and decision cycle (GFLOP); V is the data volume that must be transmitted when the complementary fraction is offloaded (Mbit); κ_{e} and κ_{c} are the edge and cloud throughputs (GFLOP s^{-1}), here 1.9 and 42, respectively; and β_{hw} is the uplink bandwidth (Mbit s^{-1}), here 8.0. Each of the three terms, therefore, carries the unit second.

And the fraction of the workload retained at the edge is chosen to minimise that total:

$$\rho^* = \arg\min_{\rho\in[0,1]} T_{\mathrm{tot}}(\rho) \tag{46}$$

Where ρ^* is the workload fraction that minimises the total response time of Equation (45) (dimensionless).

For the capacities of the environment described in Section 6.9, an edge throughput of 1.9 GFLOP per second in 15 W mode, an uplink bandwidth of 8.0 Mbit per second and a cloud throughput of 42 GFLOP per second, the solution is interior at an offload fraction of 0.82, meaning that 82% of the workload is retained at the edge. Under this partition, the per-asset uplink load is 3.6 bit per second, so a portfolio of 1,000 instrumented assets generates 3.6 kbit per second of aggregate state traffic against 45 Mbit per second if raw signals were transmitted. This four-order-of-magnitude difference, rather than any property of the models, is what makes portfolio-scale deployment tractable. The projection is arithmetic from single-asset measurements and has not been demonstrated at portfolio scale.

10.2. Deployment Barriers and Mitigation

Table 20 records the barriers to deployment, each with an assessed severity, a concrete mitigation and the risk that remains after mitigation. Gap G5 identifies the absence of this analysis as a general weakness of the field; stating the residual risk rather than only the mitigation is what makes the assessment usable by an asset owner.

Table 20. Deployment barriers, mitigation measures and residual risk

Barrier	Severity	Mitigation applied	Residual risk
Data interoperability across vendor systems	High	Mapping of all sources to an open exchange schema at ingestion; semantic alignment of asset identifiers	Legacy records without machine-readable schemas require manual mapping.
Cybersecurity of the bidirectional link	High	Mutual authentication at the edge, signed command envelopes, human authorisation for all safety-relevant actions	Supply-chain compromise of sensor firmware is not addressed by the framework.
Capital and operating cost of instrumentation	Medium	Sensor placement optimised for observability of the target state components rather than uniform coverage; 37 channels for a 60 m asset	Break-even at year 5.8 makes the case weak for assets near replacement.
Continuous power and communications	Medium	Edge nodes at 12 W with store-and-forward buffering; 72 h outage tolerated without	Assets without mains power require solar sizing, not

		reinitialisation	addressed here.
Validation under environmental uncertainty	High	Explicit normalisation, Equation (32), and uncertainty propagated through the posterior rather than discarded.	Unobserved confounders outside assumption A7 remain uncorrected.
Scalability of data volume	Medium	Edge-cloud partition of Equation (46); transmission of assimilated state rather than raw signal	Portfolio figures are arithmetic projections, not measurements.
Sensor drift, failure and replacement	Medium	Entropy weighting collapses failing channels automatically, Equation (4); verified in Section 8.4	Simultaneous failure of a channel group reduces observability.
Model validity outside the calibrated envelope	High	Residual whiteness monitoring flags divergence, Section 3.5; verified in Section 8.4	Extreme events outside assumption A3 require model reformulation.
Organisational capability and acceptance	Medium	Decision outputs expressed as inspection and intervention actions in owner terminology	Interventions triggered below visible-defect thresholds lack an accepted evidentiary basis.

10.3. Limitations and Threats to Validity

Five limitations qualify the conclusions, and they are stated without mitigation because none is available within the present study. First and most important, every result in Section 7 and Section 8 is a simulation outcome on generated data; no field validation, laboratory validation or real-time deployment has been performed, and the reported figures establish internal consistency and mechanism rather than field performance. Second, the synthetic deterioration follows a Wiener process with constant drift by construction, so assumptions A4 and A5 hold exactly in the data and cannot be tested by it; a real asset subject to sudden damage would violate them. Third, the study covers a single structural form, a three-span reinforced concrete girder bridge, so generalisation across asset classes is argued from the structure of the formulation rather than demonstrated. Fourth, the cost and carbon figures depend on labour rates, discount rate, grid carbon intensity and asset criticality specific to the assumed context and should be re-derived for any other jurisdiction. Fifth, the environmental normalisation corrects only observed confounders, so an unobserved systematic influence would remain as unexplained residual variance. The first two are threats to external validity of the strongest kind, and the priority for subsequent work is replication of Section 7 on measured data from an instrumented structure.

11. Future Research Directions

Five directions follow from the limitations above. The first and highest priority is replication on measured data: the protocol of Section 6 is designed to transfer directly to an instrumented structure, and the public benchmark literature [23, 24] provides suitable candidates. The second is standardised interoperability, since the ingestion mapping used here is bespoke, and a shared open schema for infrastructure twin exchange would remove the largest single integration cost. The third is portfolio-scale validation, extending the edge-cloud partition of Equation (46) from one asset to a network and measuring rather than projecting the resulting communication and computation load. The fourth is uncertainty-aware learned surrogates that report calibrated predictive uncertainty, which would allow the physics and data contributions to be weighted per prediction rather than by a single global coefficient of 0.62.

The fifth is cross-asset transfer, testing whether a twin calibrated on one structure accelerates convergence on a similar structure, which would materially reduce the instrumentation burden of portfolio deployment. Extension of the degradation model beyond the Wiener form, to accommodate shock loading and scour, is a necessary precondition for applying the remaining-life results to assets exposed to those hazards.

12. Conclusion

This paper has presented a unified digital twin framework for autonomous lifecycle management of civil infrastructure, integrating cyber-physical sensing, entropy-weighted multi-sensor fusion, a hybrid physics-plus-learning surrogate, extended Kalman assimilation and a reliability-constrained decision layer within a single phase-invariant state formulation.

Three contributions distinguish the framework from prior work. The phase-invariant formulation expresses design, construction, operation, maintenance and end-of-life as segments of one state trajectory, so that the posterior produced at each phase boundary serves as the prior for the next with its uncertainty intact. The blending coefficient between the physics model and the learned correction is calibrated on held-out data, obtaining 0.62 on the present dataset, rather than fixed by judgement. The decision layer enforces the failure probability as a hard constraint, rejecting the 11.7% of candidate schedules that an unconstrained objective would have accepted in breach of the target reliability. Evaluated on a provisional synthetic dataset against traditional, IoT-only and AI-only baselines under a common protocol, the framework attained 96.0% detection accuracy with a 95% Wilson interval of 95.0 to 96.8 and an area under the curve of 0.98, against 90.0% and 0.92 for the strongest baseline, with every pairwise difference significant at p below 0.001 after Holm-Bonferroni adjustment.

The remaining useful life error fell from 4.80 to 1.30 years. Median end-to-end latency fell from 250 ms to 120 ms, with a 95th percentile of 178 ms, by placing assimilation at the edge, and uplink volume fell by a factor of 1.25 by ten to the fourth.

Detection accuracy remained at or above 92% across high-temperature, heavy-load and seismic operating states. The annual failure rate fell from 0.080 to 0.020, raising twenty-year reliability from 0.202 to 0.670, with a discounted break-even at year 5.8.

The ablation attributes the margin principally to environmental normalisation at 7.4 percentage points, the physics constraint at 6.0 and continuous assimilation at 4.8, rather than to model capacity, which was held constant against the AI-only baseline. Latency reduction is shown to be architectural rather than algorithmic.

These conclusions are qualified in the strongest terms by the fact that all results are simulation outcomes on generated data: no field validation, laboratory validation or real-time deployment has been performed. Replication of the reported protocol on measured data from an instrumented structure is the necessary next step, and until it is completed, the contribution of this work is a formulation with demonstrated internal consistency rather than a validated system.

Data Integrity Statement

The status of every category of result reported in this paper is as follows. Conceptual framework outputs: the architecture of Section 3, the phase-invariant formulation of Section 4 and the application structure of Section 5, together with Eqs. (1) to (33) and Eqs. (43) to (46) are analytical constructs derived from first principles and stated assumptions; they are not empirical claims. Provisionally generated data: the dataset described in Section 6.1 to Section 6.3, comprising Tables 6, 7 and 8, is synthetic.

It was generated deterministically from the mechanical relationships stated in Section 6.2 for the purpose of demonstrating and testing the framework. It is not field-measured, not laboratory-measured and not sensor-recorded. Simulated validation results: every quantitative result in Sections 7 and 8, comprising Tables 10 to 20 and Figures 1 to 5, is a simulation outcome computed on that synthetic dataset.

Literature-derived benchmark ranges: the baseline modal frequencies, damping ratios, thermal sensitivities, sensor noise levels and comparative accuracy ranges cited in Section 6.1, Table 6 and Section 9.6 are bounded by published values for comparable structures and are attributed at the point of use. Measured or experimentally verified results: none.

No field validation, laboratory validation, real-time deployment or physical measurement has been performed in connection with this work, and no statement in this paper should be read as claiming otherwise. The dataset generator, configuration files and random seeds are released so that every reported figure can be reproduced exactly.

Funding Statement

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Conflicts of Interest

The authors declare no conflict of interest. No citation in this manuscript has been included other than on the grounds of its technical relevance to the work reported. No author self-citation and no citation to the publishing journal has been added for any purpose other than direct technical relevance.

Data Availability Statement

All data supporting the findings of this study are synthetic and are fully reproducible from the released materials. The dataset generator implementing Tables 6 and 7, the preprocessing and model code, the exact partition indices, the configuration files, the five random seeds and the trained model weights are deposited in a public repository under a permissive research licence; the persistent identifier will be supplied at acceptance and quoted in the published version. No proprietary, restricted or personally identifying data were used in this study.

Acknowledgements

The authors thank the reviewers of the International Journal of Civil Engineering, whose comments on methodological completeness, statistical rigour and reproducibility prompted the substantial revision of which this manuscript is the result.

Mathematical Notation and Symbol Conventions

Notation follows a single convention throughout. Scalar variables are set in italic; vectors and matrices in bold italic; and units, abbreviations, numerals, operators and standard mathematical functions upright. Subscripts that denote an abbreviation, such as for threshold or phy for physics, are upright, whereas subscripts that are running indices are italic. Every symbol is defined at the point of first use, together with its unit.

References

[1] Shi Qiu et al., "Exploring the Impact of Digital Twin Technology in Infrastructure Management: A Comprehensive Review," *Journal of Civil Engineering and Management*, vol. 31, no. 4, pp. 395-417, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[2] Chongjie Kang et al., "Digital Twin Technologies for Bridge Lifecycle Management—Literature Insights and a Pilot Study on the Nibelungen Bridge," *Results in Engineering*, vol. 28, pp. 1-19, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[3] Hessam Kaveh, and Reda Alhajj, "Advancing Civil Infrastructure with Digital Twins: A Review of Applications and Challenges," *Journal of Civil Engineering and Management*, vol. 31, no. 8, pp. 828-842, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[4] Ahmad Baghdadi, "A Comprehensive Review of Digital Twin Implementation in Construction: Current Trends and Future Directions," Construction Innovation," *Journal of Asian Architecture and Building Engineering*, vol. 25, no. 4, pp. 3622-3636, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[5] Helly Kamdar, "Digital Twins and Automation in Construction Project Lifecycle Management," *Emerging Engineering and Mathematics*, vol. 3, no. 1, pp. 74-83, 2025. [Publisher Link]

[6] Zhiyan Sun et al., "Approach Towards the Development of Digital Twin for Structural Health Monitoring of Civil Infrastructure: A Comprehensive Review," *Sensors*, vol. 25, no. 1, pp. 1-41, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[7] Hossein Naderi, and Alireza Shojaei, "Digital twin technologies for civil infrastructure: Current state and future directions," Automation in Construction, vol. 156, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[8] Fei Tao et al., "Digital Twin in Industry: State-of-the-Art," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 4, pp. 2405-2415, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[9] Rui Zhang et al., "Digital Twin and its Applications: A Survey," *The International Journal of Advanced Manufacturing Technology*, vol. 123, no. 11-12, pp. 4123-4136, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[10] Adam Thelen et al., "A Comprehensive Review of Digital Twin - Part 1: Modeling and Twinning Enabling Technologies," *Structural and Multidisciplinary Optimization*, vol. 65, no. 12, pp. 1-75, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[11] Michael G. Kapteyn, Jacob V. R. Pretorius, and Karen E. Willcox, "A Probabilistic Graphical Model Foundation for Enabling Predictive Digital Twins at Scale," *Nature Computational Science*, vol. 1, no. 5, pp. 337-347, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[12] Feng Jiang et al., "Digital Twin and its Implementations in the Civil Engineering Sector," *Automation in Construction*, vol. 130, pp. 1-38, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[13] Werner Kritzinger et al., "Digital Twin in Manufacturing: A Categorical Literature Review and Classification," *IFAC-PapersOnLine*, vol. 51, no. 11, pp. 1016-1022, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[14] Song Honghong et al., "Digital Twin Enhanced BIM to Shape Full Life Cycle Digital Transformation for Bridge Engineering," *Automation in Construction*, vol. 147, pp. 1-40, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[15] M. Pregnolato et al., "Towards Civil Engineering 4.0: Concept, Workflow and Application of Digital Twins for Existing Infrastructure," *Automation in Construction*, vol. 141, pp. 1-15, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[16] Haidar Hosamo Hosamo et al., "A Review of the Digital Twin Technology in the AEC-FM Industry," *Advances in Civil Engineering*, vol. 2022, pp. 1-17, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[17] Itxaro Errandonea, Sergio Beltrán, and Saioa Arrizabalaga, "Digital Twin for Maintenance: A Literature Review," *Computers in Industry*, vol. 123, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[18] Deuk-Young Jeong et al., "Digital Twin: Technology Evolution Stages and Implementation Layers with Technology Elements," *IEEE Access*, vol. 10, pp. 52609-52620, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[19] Michael Grieves, and John Vickers, "Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems," *Transdisciplinary Perspectives on Complex Systems*, pp. 85-113, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[20] Louise Wright, and Stuart Davidson, "How to Tell the Difference between a Model and a Digital Twin," *Advanced Modeling and Simulation in Engineering Sciences*, vol. 7, no. 1, pp. 1-13, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[21] Yuanliang Yang et al., "Research Progress and Prospect of Digital Twin in Bridge Engineering," *Advances in Structural Engineering*, vol. 27, no. 2, pp. 333-352, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[22] Samad M. E. Sepasgozar, "Differentiating Digital Twin from Digital Shadow: Elucidating a Paradigm Shift to Expedite a Smart, Sustainable Built Environment," *Buildings*, vol. 11, no. 4, pp. 1-16, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[23] Marios Impraimakis, and Evangelia Nektaria Palkanoglou, "A Generative Adversarial Network Optimization Method for Damage Detection and Digital Twinning by Deep AI Fault Learning: Z24 Bridge Structural Health Monitoring Benchmark Validation," *Structural and Multidisciplinary Optimization*, vol. 68, no. 11, pp. 1-21, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[24] Stefan Wernitz et al., "A New Open-Database Benchmark Structure for Vibration-based Structural Health Monitoring," *Structural Control and Health Monitoring*, vol. 29, no. 11, pp. 1-26, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[25] Jabez Nesackon Abraham et al., "Unsupervised Learning-Based Anomaly Detection for Bridge Structural Health Monitoring: Identifying Deviations from Normal Structural Behaviour," *Sensors*, vol. 26, no. 2, pp. 1-20, 2026. [CrossRef] [Google Scholar] [Publisher Link]

[26] Qiuchen Lu et al., "Digital Twin-Enabled Anomaly Detection for Built Asset Monitoring in Operation and Maintenance," *Automation in Construction*, vol. 118, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[27] Kay Rogage et al., "Beyond Digital Shadows: Digital Twin used for Monitoring Earthwork Operation in Large Infrastructure Projects," *AI in Civil Engineering*, vol. 1, no. 1, pp. 1-21, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[28] Linjun Lu et al., "Modeling Heterogeneous Spatiotemporal Pavement Data for Condition Prediction and Preventive Maintenance in Digital Twin-Enabled Highway Management," *Automation in Construction*, vol. 174, 1-19, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[29] Diana Davletshina, Varun Kumar Reja, and Ioannis Brilakis, "Automating Construction of Road Digital Twin Geometry using Context and Location Aware Segmentation," *Automation in Construction*, vol. 168, pp.1-17, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[30] Somin Park et al., "Natural Language Instructions for Intuitive Human Interaction with Robotic Assistants in Field Construction Work," *Automation in Construction*, vol. 161, pp. 1-57, 2024. [CrossRef] [Google Scholar] [Publisher Link]