Data Engineering and Applications

Research Paper

# Physics-Guided Graph Network-Driven State Estimation Model for Bridge Digital Twins

**Murat Öztürk[1] and Selim Aydın[1*]**

[1] Faculty of Engineering, Bilkent University, Ankara, 06800, Turkey
*Corresponding author: selim.ay@bilkent.edu.tr

**Abstract.** If the measurements are sparse, asynchronous, noisy, and influenced by traffic and temperature, it is not possible to continuously synchronize the physical bridge with its digital counterpart. While unconstrained deep estimators may fit the observed channels but violate equilibrium or yield implausible unmeasured responses, pure finite-element updating is interpretable but computationally limited. In this paper, we develop a graph state-space network for bridge digital-twin state estimation guided by physics. The bridge is depicted as a sensor-component graph with numerous accelerations, strain, displacement, temperature, and traffic variables as nodes. After modal displacement and velocity are predicted by a reduced structural transition operator, observations along physically significant edges are incorporated into a gated graph correction block. Equilibrium, observation, smoothness, and uncertainty-calibration losses stabilize joint state and stiffness-change estimation. Predictive covariance is propagated to the digital twin by a streaming synchronization technique that uses mask-conditioned messages to handle missing packets. Long-span bridge monitoring replay, a laboratory steel-girder test, and a calibrated finite-element benchmark with traffic and thermal stress have all been employed in the experiments. When compared to strong recurrent, graph, and Kalman baselines, the suggested model lowers the displacement root-mean-square error by 24.8%–41.6% in all three scenarios. The normalized state error is 0.086 under 30% random sensor loss, and the F1 score for stiffness-loss localization is 0.914. The model uses a 64-sample window of 18.7 ms per GPU, with an empirically covered interval of 92.8% at the nominal 95% level. According to the findings, incorporating succinct mechanics into topology-aware learning enhances real-time twin operation's accuracy, robustness, and diagnostic reliability.

**Keywords:** *Bridge Digital Twin, Physics-guided Graph Network, State-space Estimation, Structural Health Monitoring*

Received on 13 January2023, Accepted on 23 April 2023, Published on 06 May 2023



## Introduction

Periodic inspections will give way to ongoing digital recordings of geometry, mechanics, monitoring data, and other information. Three-dimensional information models can organize condition evidence and support maintenance decisions, as the first bridge twins showed [1]. Visual correspondence is not the foundation of their practical applications. In addition to other values that cannot be directly measured and parameters that gradually change as a result of aging, temperature, or cumulative load, the virtual bridge must be updated to reflect any changes to the actual structure. Although they have produced extensive time-series data, modern structural health monitoring systems frequently lack full-space observation. While cable length is not an issue for wireless networks, they do have problems with clock drift, packet loss, energy limits, and heterogeneous noise [2]. Acceleration, strain, displacement, temperature, and traffic systems all see distinct projections of the structural condition; optical measures can be utilized to assess the dispersed motion without contact and have demonstrated satisfactory results in seismic response evaluation [3]. Therefore, the computational foundation that separates an operational digital twin from a static digital model is state estimation.

Data-centric monitoring has improved the screening and compression of anomalies for long-span bridges [4], and laser scanning and geometric reconstruction have expanded the information available to bridge twins [5].

Dynamic synchronization is not addressed by the aforementioned developments. An accurate signal predictor will not be able to estimate a response that fits the real mass, damping, boundary conditions, or load routes, and a geometry-rich model can still retain the outdated stiffness values. The need for continuous lifecycle data is highlighted by fatigue-oriented twin frameworks [6], and the need for simultaneous consideration of both local and global responses is demonstrated by hybrid monitoring of orthotropic decks [7]. When there are unfavorable environmental changes, a lot of traffic, or partial sensor failures, the estimate problem will be more severe. These regimes can be described by high-fidelity finite-element models, but they remain susceptible to ambiguous boundary conditions and are too costly for millisecond assimilation. Although classical Kalman filters are straightforward and efficient, they are not appropriate for nonstationary residual behavior because they assume linear dynamics and constant noise. Although their hidden states are often unstructured, sequence networks are able to learn non-linear temporal patterns. A good compromise is provided by graph networks, which allow for the explicit encoding of bridge topology, sensor layout, and component adjacency instead of flattening them into an arbitrary vector. Multimedia and knowledge representations have also been used by current digital-twin monitoring systems to link different bridge information [8]. However, there is currently no unified estimator that incorporates interpretable states, structural message passing, fault adaptation, stiffness-change estimation, and calibrated uncertainty.

In order to overcome the aforementioned shortcomings, this research develops a graph network-based bridge digital twin that is informed by physics. The modal displacement and velocity are first explicitly predicted via a reduced-order structural transition. A mask-conditioned graph correction preserves bridge connectivity while learning the residual of imperfect mechanics and multimodal observations. Component stiffness ratios and predictive uncertainty are jointly estimated by a dual state-parameter head, and the estimator outputs are transformed into versioned twin states appropriate for replay and maintenance queries by a synchronization service. State correctness, fault tolerance, damage localization, uncertainty calibration, and computational delay are among the evaluation metrics. In the remaining sections of this study, related techniques are introduced, the estimate problem is formulated, the suggested design and implementation are shown, and controlled and field-replay experiments are reported.

## Related Work

### Graph Learning for Structural Monitoring

Because sensors have members, boundary conditions, modal participation, and load paths, structural monitoring data are inherently relational. These relationships can be used formally in graph signal processing and graph neural networks, and their usefulness in monitoring tasks has been demonstrated by a computational framework for complex structural sensor networks [9]. Unlike one-dimensional convolutions, graph operators can maintain the non-uniform layout of sensors. It has constructed connections and only learns small adaptive variations; thus, it is not free-form self-attention.

Additionally, graph models have been used to use spatiotemporal attention to reconstruct the missing structural monitoring sequence [10]. Earlier convolutional recovery methods learned local signal regularities [11], and group-sparsity-aware networks explicitly handled continuous missing blocks [12]. The aforementioned methods can use space redundancy to recover missing data, however they typically focus on signal completion rather than latent structural state estimation. When unmeasured displacement, velocity, internal modal coordinates, or a change in component stiffness are the required output, reconstructing a decent sensor trace is insufficient.

For monitoring structural health, Deep Residual Networks enhanced feature learning [13]. Inadequate data quality also affects model reliability for bridge deployment. For bridge monitoring streams, frequency-informed deep evaluation has been proposed [14], and statistical plus deep-learning techniques have identified inaccurate bridge observations [15]. The graph node characteristics' explicit measurement masks and quality indicators are motivated by these contributions. They also point out a weakness: rather than being a factor influencing the assimilation gain in the state estimator, data quality is frequently handled as a pre-processing step.

**Digital Twins and Physics-Guided Estimation**

Backpropagation networks and recurrent memory have been employed to represent degradation trends in long-term bridge health monitoring [16]. Identification of beam-bridge deterioration has been aided by recurrence-graph representations [17]. The aforementioned techniques typically use training examples to indirectly extract valuable patterns. If sensor availability or operational conditions differ from those in the training data, generalization may be diminished. Another method of incorporating known dynamics into learning is through physics-informed deep models. Physical guidance can limit degradation inference, as demonstrated by multisensory bearing prognosis [18]. A complete partial differential equation solver is not always the most useful type of physics guidance for bridges. Inertia, damping, stiffness, and forcing are already included in a low-order modal model in coordinates that are consistent with streaming estimation. A graph network can then learn the remaining discrepancy, which is caused by truncation, imperfect loads, nonlinear joints, and temperature-dependent boundaries.

As a result, the literature has produced three interrelated flaws. First, state estimators are needed for digital twins rather than only response prediction or data reconstruction. Second, graph learning should provide messages with an assimilation function by operating within an explicit predict-correct cycle. Third, a confident but mechanically inconsistent estimate will result in an incorrect maintenance decision; therefore, uncertainty and stiffness change should be estimated concurrently with dynamic states. In addition to addressing these shortcomings, the new framework is small enough to be deployed online.

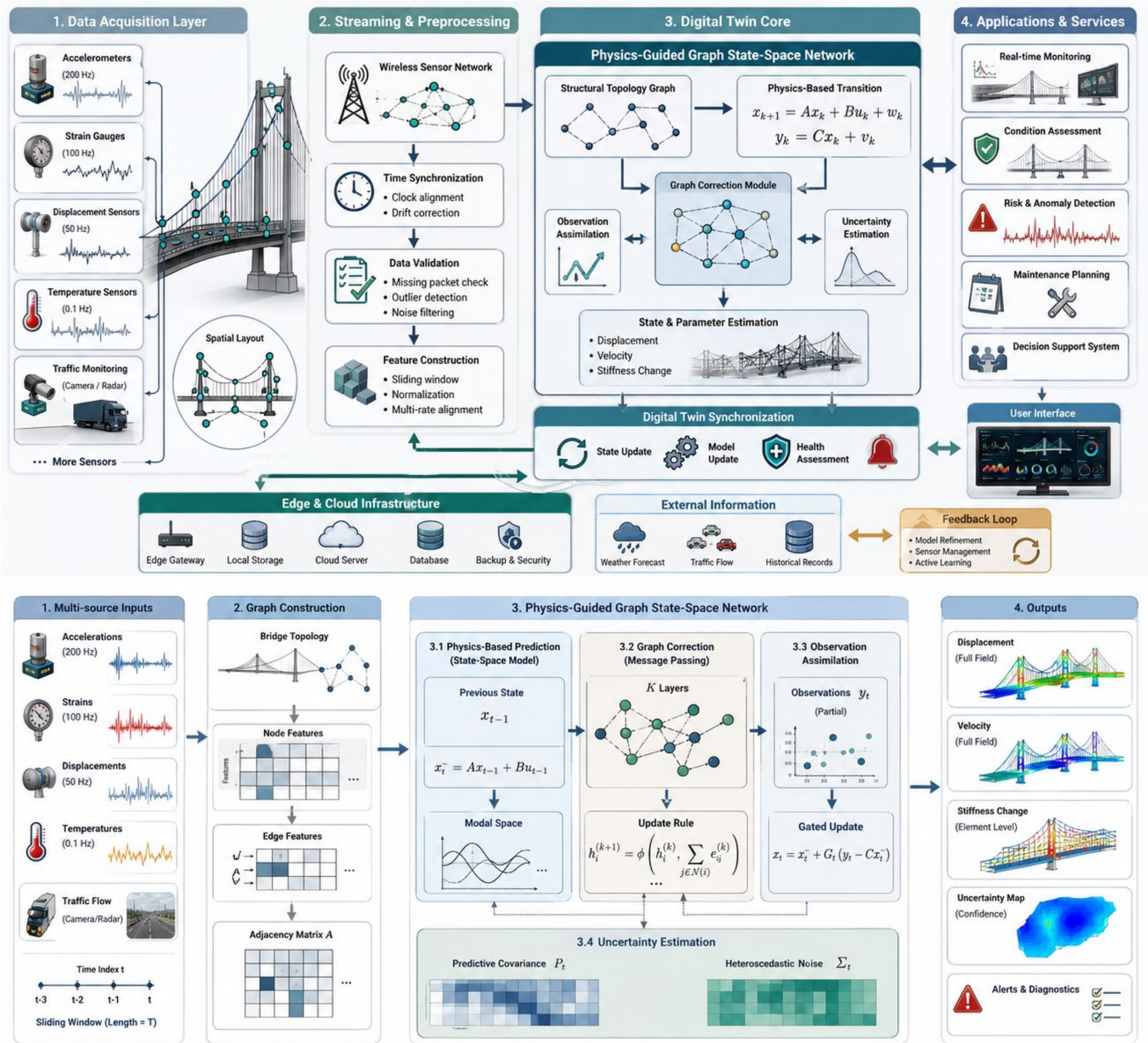



**Figure 1.** Conceptual information flow of the physics-guided graph bridge digital twin.

## Bridge Digital-Twin State Estimation Problem

### State, Observation, and Graph Definition

The physical bridge's parts are mechanically coherent or instrumented. The edges are shared members, finite-element connection, modal correlation, or validated sensing dependencies, whereas each graph node represents a sensing site or a condensed structural degree of freedom. A useful basis for this digital-twin reduction is provided by low-fidelity physical guiding [19]. In addition to geometry, reduced matrices, sensor metadata, and the most recent posterior state, Twin stores a time-stamped graph. The information flow from physical acquisition to mechanics prediction, graph rectification, uncertainty evaluation, and twin-state publication is depicted in Figure 1.

Let the latent vector contain retained modal displacement, modal velocity, and component stiffness ratios. The measured vector contains synchronized acceleration, strain, displacement, temperature, and traffic descriptors after unit normalization. A binary availability mask and a continuous quality score accompany every measurement. This separation is important: an absent channel contributes no innovation, whereas a low-quality channel may contribute weakly if neighboring sensors and the physical prior agree.

$$\mathrm{x}_t = [\mathrm{q}_t, \dot{\mathrm{q}}_t, \mathrm{\rho}_t]^{\mathrm{T}} \quad \text{Eq.(1)}$$

The retained modal coordinates and velocities at time $t$ are $\mathrm{q}_t$ and $\dot{\mathrm{q}}_t$, respectively, while $\mathrm{\rho}_t$ contains stiffness ratios for predefined bridge components. A ratio of one denotes the reference condition; a sustained decrease suggests stiffness loss. Keeping the vector compact allows real-time inference while maintaining variables that can be mapped back to physical displacement, strain, and component condition.

$$G = (V, E, A) \quad \text{Eq.(2)}$$

The graph is defined by its node set $V$, edge set $E$, and weighted adjacency matrix $A$. Initial weights decay with physical distance and are increased for nodes sharing strong modal participation. During training, a bounded adaptive residual refines $A$, but a penalty prevents the learned graph from erasing the known load-path structure.

### Mechanics-Based Transition and Observation

The reduced structural model is obtained from a calibrated finite-element baseline. Modal truncation retains the modes needed to reproduce 95% of measured kinetic energy over the monitoring band. Temperature-dependent boundary effects are represented by a small affine correction, while unmeasured traffic excitation is summarized by an inferred generalized force. The estimator advances the state using a discrete transition before any learned correction is applied.

$$\bar{\mathrm{x}}_t = F(\mathrm{\rho}_{t-1}, T_t)\hat{\mathrm{x}}_{t-1} + B\mathrm{f}_t \quad \text{Eq.(3)}$$

The prior state $\bar{\mathrm{x}}_t$ is produced from the previous posterior $\hat{\mathrm{x}}_{t-1}$. The matrix $F$ is the discretized modal transition conditioned on the stiffness ratios and temperature $T_t$, while $B$ maps the estimated generalized force $\mathrm{f}_t$. The overbar identifies a prior quantity. This calculation supplies inertia and continuity even when all sensors are temporarily unavailable.

$$\bar{\mathrm{y}}_t = H\bar{\mathrm{x}}_t \quad \text{Eq.(4)}$$

The predicted measurement $\bar{\mathrm{y}}_t$ is obtained by mapping the prior state through $H$. This operator includes modal-shape interpolation and the appropriate derivative or strain transformation for each sensor type. The innovation is formed only after time alignment and normalization, so its magnitude remains comparable across heterogeneous channels.

### Estimation Objective and Operational Constraints

The operational objective is to minimize state error while preventing physically impossible corrections. Since true states are available only in simulation and controlled laboratory tests, training mixes supervised windows with self-supervised field windows. The estimator must also remain stable under burst loss, timestamp jitter, sensor bias, and loads outside the calibration range. These constraints are evaluated separately rather than absorbed into a single average score.

$$\mathrm{r}_t = \mathrm{m}_t \odot (\mathrm{y}_t - \bar{\mathrm{y}}_t) \quad \text{Eq.(5)}$$

https://doi.org/10.64972/dea.2023.v2i2.3564d:46-62

The masked innovation $\mathrm{r}_t$ combines the binary availability vector $\mathrm{m}_t$ with the synchronized observation $\mathbf{y}_t$ and predicted measurement $\bar{y}_t$ through elementwise multiplication $\odot$. Quality scores are not included in the mask; they enter the graph gates so the network can attenuate rather than abruptly discard a questionable measurement.

The resulting problem is a dual estimation task. Dynamic states vary at the sampling rate, whereas stiffness ratios should evolve slowly unless a structural event occurs. A shared graph encoder processes the innovation, but separate heads and temporal regularizes enforce these different time scales. Alarm logic is applied only after posterior estimates persist across multiple windows, reducing sensitivity to isolated disturbances.

Observability is assessed before learning. For each candidate sensor configuration, the reduced observability Grampian is evaluated over the 64-sample assimilation horizon, and configurations with poorly excited retained modes are flagged. The graph network can interpolate information from neighboring nodes, but it cannot create evidence for a mode that is absent from both the measurements and the transition. This check prevents misleading confidence when a nominally dense network contains sensors that are redundant with respect to the states of interest. It also defines which reconstructed quantities may be published as estimates and which should remain unavailable.

The estimator distinguishes epistemic model discrepancy from ordinary measurement noise only operationally, not as a mathematically exact decomposition. Automated structural clones have likewise emphasized the need to accommodate nonideal connection behavior [20]. Observation variance is initialized from calibration records and updated slowly, whereas the learned covariance responds at window scale. Persistent equilibrium residual, broad graph-gate disagreement, and increasing posterior variance jointly indicate that the twin is outside its supported domain. This diagnostic state blocks automatic parameter updates but continues to publish dynamic estimates with an explicit degraded-quality flag.

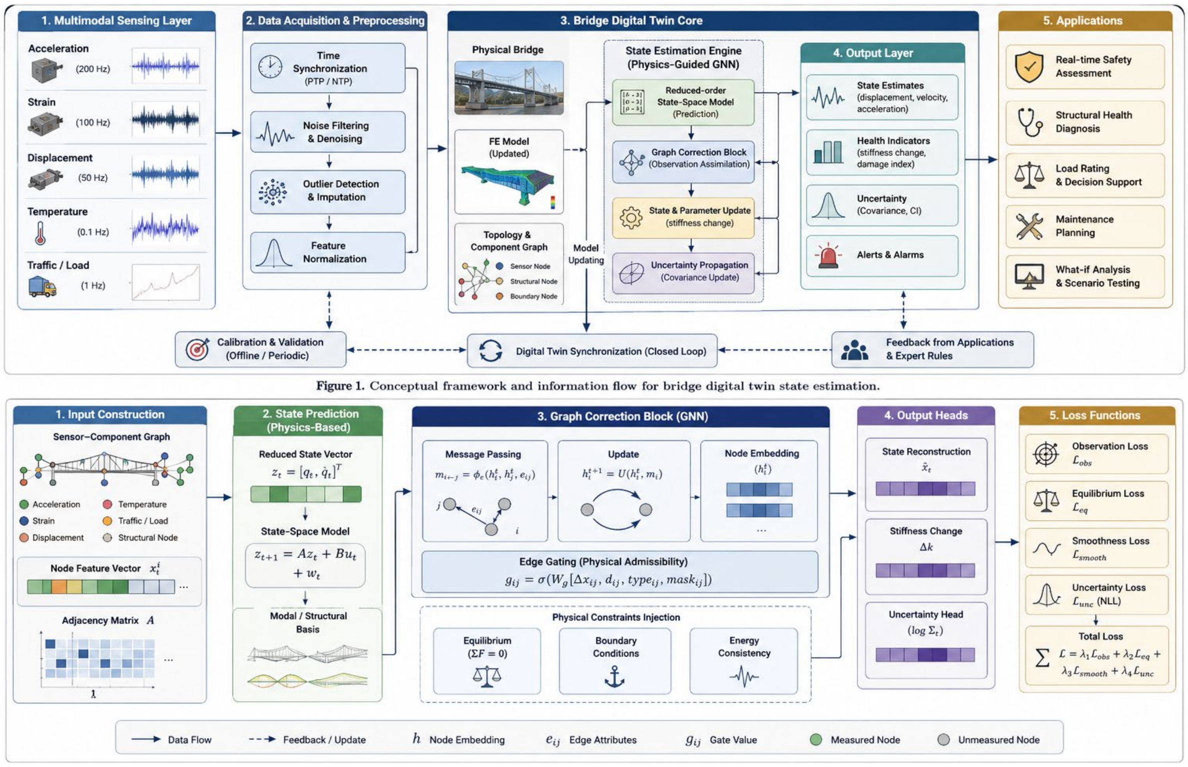



**Figure 2.** Architecture of the mechanics predictor, gated graph correction, and dual state–parameter heads.

## Physics-Guided Graph State-Space Network

### Topology-Aware Innovation Encoding

The suggested model architecture is shown in Figure 2 and consists of the mechanics predictor, dual state-parameter outputs, observation assimilation, topology-aware graph rectification, and uncertainty estimation. During correction, Bayesian monitoring encourages displaying the uncertainty at all times [21]. Predicted and observed measurements, innovation, availability mask, quality score, temperature, sensor type embedding, and local modal participation are all sent to each node. Physical distance, member continuity, common-mode involvement, and data-driven correlation calculated exclusively during the training phase are all encoded by edge features.

Messages are computed by a two-layer graph network with 64 hidden units. A source message is transformed jointly with the edge descriptor, then scaled by a gate derived from source quality, target quality, and innovation consistency. This construction prevents a faulty high-amplitude channel from dominating its neighborhood. Residual connections preserve local evidence, and layer normalization controls scale drift across sensor types.

$$\mathrm{h}_i^{k+1} = \sigma\left(W_0 \mathrm{h}_i^k + \sum_{j \in \mathcal{N}_i} \alpha_{ij} W_1 \mathrm{h}_j^k\right) \quad \text{Eq.(6)}$$

At graph layer $k$, the hidden feature $\mathrm{h}_i^{k+1}$ is updated using learned transformations $W_0$ and $W_1$, the physical neighborhood $\mathcal{N}_i$, the smooth activation $\sigma$, and the quality-aware edge gate $\alpha_{ij}$. The sum is degree-normalized in implementation, which avoids systematically larger corrections at highly connected nodes.

$$\alpha_{ij} = \mathrm{sigmoid}\left(\mathrm{a}^{\mathsf{T}}\left[\mathrm{h}_i, \mathrm{h}_j, \mathrm{e}_{ij}, s_i, s_j\right]\right) \quad \text{Eq.(7)}$$

The edge gate $\alpha_{ij}$ is computed from the source and target features $\mathrm{h}_i$ and $\mathrm{h}_j$, the edge descriptor $\mathrm{e}_{ij}$, and the quality scores $s_i$ and $s_j$. The gate remains interpretable as a bounded assimilation weight. Its average value can be logged by sensor pair, giving operators a trace of which observations influenced a twin update.

### Physics-Guided Correction and Dual Estimation

The pooled graph representation does not replace the structural prior. Instead, node corrections are projected to modal space and added to the prior through a learned gain. The gain is bounded with a hyperbolic tangent and scaled by the prior covariance, making large corrections possible only when the model is uncertain or the innovation is spatially coherent. A separate parameter head aggregates slower features over 30 windows to estimate stiffness ratios.

$$\hat{\mathrm{x}}_t = \bar{\mathrm{x}}_t + K_\theta\left(H_G(\mathrm{r}_t, A)\right) \quad \text{Eq.(8)}$$

The graph encoder $H_G$ and the learned bounded correction map $K_\theta$ transform the innovation into the posterior state $\hat{\mathrm{x}}_t$ while keeping the estimate anchored to the mechanics prediction. When measurements disappear, the encoded innovation approaches zero and the update naturally falls back toward the structural transition rather than hallucinating a measurement-driven state.

$$\hat{\rho}_t = \mathrm{clip}(\hat{\rho}_{t-1} + \mathit{\Delta}\rho_t, 0.6, 1.1) \quad \text{Eq.(9)}$$

The slow parameter increment $\mathit{\Delta}\rho_t$ updates the stiffness ratio, which is clipped to a broad engineering range for training stability. The limits are not alarm thresholds. Maintenance alerts use posterior probability, spatial persistence, and agreement with strain residuals, so the estimator does not convert every bounded change into a damage declaration.

### Loss Function and Uncertainty Propagation

Training minimizes a combination of state supervision, observation reconstruction, equilibrium consistency, stiffness smoothness, and probabilistic calibration. Supervised state terms are active for simulated and laboratory windows. Field replay relies on observation and physics terms, with pseudo-labels used only when the calibrated baseline and multiple sensor families agree. Noise augmentation covers Gaussian noise, colored drift, bias steps, timestamp jitter, and contiguous missing blocks.

$$\mathcal{L} = \mathcal{L}_{\text{state}} + \lambda_y \mathcal{L}_y + \lambda_p \mathcal{L}_p + \lambda_s \mathcal{L}_s + \lambda_u \mathcal{L}_u \quad \text{Eq.(10)}$$

The total objective combines five concise losses. $\mathcal{L}_{\text{state}}$ measures normalized state error, $\mathcal{L}_y$ reconstructs available observations, $\mathcal{L}_p$ penalizes the reduced-equilibrium residual, $\mathcal{L}_s$ discourages rapid stiffness variation, and $\mathcal{L}_u$ is a Gaussian negative log-likelihood for uncertainty calibration. The weights are selected on the validation set using state accuracy as the primary criterion and interval calibration as a feasibility constraint.

$$\hat{P}_t = \mathrm{diag}\big(\mathrm{softplus}(\mathrm{z}_t)\big) + \varepsilon I \qquad \text{Eq.(11)}$$

The uncertainty-head output $\mathrm{z}_t$ is converted into the positive posterior diagonal covariance $\hat{P}_t$. The small term $\varepsilon I$ prevents numerical collapse. Although cross-covariance is not explicitly predicted, spatial dependence enters through the graph encoder; empirical coverage and calibration error are reported to expose any remaining overconfidence.

The architecture contains 0.84 million trainable parameters. This scale is deliberate: the reduced transition handles predictable dynamics, leaving the network to model discrepancy and assimilation. A larger transformer provided no consistent validation gain and increased latency, while a smaller graph encoder underfit cross-girder interactions during asymmetric traffic. The selected model therefore balances representational capacity with edge-deployment constraints.

Several safeguards constrain the learned correction. Event-triggered estimation supports selective transmission of informative innovations [22]. Modal increments are scaled by validation-set standard deviations, and a spectral-radius check is applied to the locally linearized closed-loop update on sampled windows. If the estimated radius exceeds one for three consecutive windows, the service attenuates the graph correction and reverts toward the mechanics prediction. Gradient clipping at a norm of five prevents a small number of extreme simulated loads from defining the update map.

Interpretability is provided at three levels. Nonlinear state-estimation studies establish sensor placement as an observability problem rather than a coverage-only problem [23]. Node gates identify which sensors supported a correction, the modal projection reveals which global response patterns changed, and the parameter head maps persistent discrepancies to predefined components. The system does not claim that an attention-like weight is a causal damage explanation. Instead, it provides a reproducible influence trace that can be compared with residual sign, sensor quality, and structural mode shape.

## Experimental Results

### Model Preparation, Training, and Evaluation Setup

First, the bridge's finite-element representation has been used to derive a reduced model. The baseline matrices serve as a stable mode, an approximate force path, and an interpretable transition for the graph residual to operate on; they are not thought to be exact. Rather of being hidden by forceful calibration, boundary and connection inconsistencies are kept for taught rectification.

Rather than being expanded into an expensive nonlinear solver, boundary and connection nonlinearities are summarized as discrepancy features. A brief alert that the linear reduced transition is departing the trusted region is provided by residual-energy indicators and temperature interactions.

Calibration coefficients, orientation, units, sampling rate, clock source, component assignment, and versioning are all included in sensor information. The pipeline maintains a parallel 1 Hz stream of temperature and parameter trends while resampling all channels to 100 Hz for dynamic estimation. Only high-value innovations are chosen for transmission, and uncertainty is incorporated into the decision-making process.

There is a limited-latency buffer in synchronization. Later packets are written to the replay storage without altering the previously published online state; packets arriving within 200 ms are aligned by the revised timestamp. In field replay, an edge device reduces average uplink traffic by 57% by sending full windows following an innovation-energy trigger and compact statistics otherwise.

Set up the graph according to sensor placement and finite-element connectivity. Candidate sensors are assessed using modal observability, redundancy, and access constraints. Instead of being a straightforward coverage issue, placement is now seen as an estimating design problem. In training, adjacency adaptation is restricted to 10% of the physical weight while maintaining the graph's engineering interpretation.

There is a 50% overlap and 64 samples in the training windows. Vehicle speeds between 10 and 35 m/s, axle groups between 80 and 520 ken, temperature changes between -15 and 35 °C in relation to calibration, damping ratios between 0.8% and 2.5%, and stiffness reductions between 2% and 25% in one or two components are all included in the simulated data. Explicit coverage of the rare response tail is driven by extreme-stress prediction [24]. The adoption of globally meaningful displacement targets is supported by studies on random traffic deflection [25].

Reducing traffic-induced deflection is also another goal. 48,000 simulated windows, 7,200 laboratory windows, and 18,600 unlabeled field windows make up the training set. To avoid temporal leakage, the validation and test partitions are divided by loading run or calendar week.

During training, sparse layouts are displayed as contiguous region masking and node dropout. Useful condition information can endure extreme instrumentation economy, as demonstrated by single-sensor bridge monitoring [26]. After a maintenance outage, the same twin service can be used because the model was trained on the layouts and does not require retraining for each missing sensor.

Strain and inclination can be used with acceleration if displacement channels are unavailable. Instead, then using test-set statistics, Channel Normalization makes advantage of the training set's robust median and interquartile range statistics.

When traffic descriptors are available, the force encoder uses them; otherwise, structure observation can be used to infer a general force. A modality dropout probability of 0.15 is set to prevent dependence on a single traffic subsystem, and bias augmentation is performed after normalization to keep sensor drift visible to the quality gate.

Three interfaces are included in the streaming service for the trained estimator: window ingestion, posterior-state publication, and deterministic replay. Temporary mobile nodes are accepted by the interface, but their graph edges have poor initial quality and must repeatedly agree with fixed sensors to pass additional gates.

Modal coordinates, reconstructed component responses, stiffness ratios, uncertainty, data-quality flags, and provenance hashes are published as the twin state. Traffic vibrations are not the same as slow hazards. As a result, the service offers both 30-minute parameter summary and 0.64-second dynamic windows.

Stiffness ratio is not the only factor used to determine damage. Stiffness probability, graph-local strain residual, modal-frequency drift, and persistence are all included in the service. Only after three indicators surpass their validation thresholds for five consecutive summary periods will a component be assigned for engineering review.

Fixed data divisions, recorded random seeds, unit-tested preprocessing, exported graph versions, and a complete replay manifest all ensure reproducibility. Adam was used for 120 epochs of training with a starting learning rate of $10^{-3}$, cosine decay, weight decay of $10^{-1}$, and early halting after 15 validation epochs. The deployment latency is measured following a 200-window warm-up, and all published results are the average of five seeds. There is a staged deployment process in place. The service does not write condition parameters to the authoritative twin during the shadow operation; instead, it estimates states. The sensor graph is approved after high-residual windows are examined and the reconstructed response is compared to the withheld channel. Alarms stay advisory until at least one controlled load event validates the response scale after parameter writing is turned on with a conservative rate limit. This procedure will give a clean baseline for further drift analysis and prevent an early calibration error from being carried over into the maintenance record.

Replay and Data Governance Together. Preprocessing version, input-window identifier, reduced-matrix checksum, graph checksum, posterior carry model version, and quality decisions applied to each channel. As a result, a subsequent calibration modification will generate a new state branch rather than replacing the earlier estimate. The same acquisition and estimation environment can be used by operators to replicate an alarm. Although this provenance is computationally straightforward, it is required if the twin output is used to compare the behavior of the structure before and after maintenance or to support the choice of inspection priority.

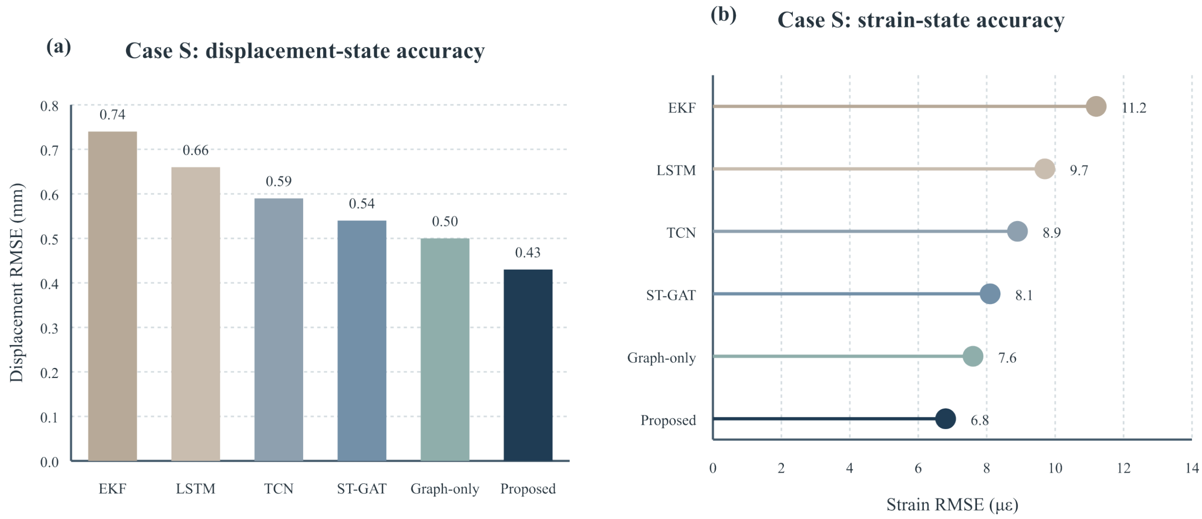
**State Accuracy, Robustness, and Ablation**

We used three corresponding situations. With 96 sensing nodes, 14 retained modes, changing traffic, thermal drift, and known component stiffness, Case S is a calibrated finite-element cable-stayed bridge. Case L is a 12.0-meter laboratory steel-girder bridge that has two displacement transducers, 16 strain gauges, 24 accelerometers,

https://doi.org/10.64972/dea.2023.v2i2.3564d:46-62

and detachable bolted-plate stiffness adjustments of 5%, 10%, and 20%. Case F is evaluated using held-out sensor reconstruction, modal consistency, event agreement, and expert-screened anomaly intervals because its internal state is not fully observable. Case F replays 14 months of long-span bridge monitoring from 68 channels. Baselines include a temporal convolutional network, a spatiotemporal graph attention network, a graph network without physics loss, an extended Kalman filter using the same reduced model, and a long short-term memory estimator. The windows, masks, and training partitions of every learnt baseline are identical. The same 40-trial budget is used to optimize hyperparameters. Displacement and Strain Root-Mean-Square Error, normalized state error, stiffness localization F1, 95% interval coverage, processing latency, and expected calibration error are the metrics.

The primary tests employ full sensors and include contiguous regional losses, noise amplification, temperature extrapolation, unobserved traffic speeds, and random loss rates ranging from 10 to 50%. Limited spatial contrast can nonetheless facilitate bridge-condition assessment under moving loads, according to two-sensor monitoring investigations [27]. One injured site will be used for the test in the lab. The last three months have been set aside by the field replay for a week of high temperatures and a sensor maintenance outage. Operationalization and interpolation are two different things.




**(c)Cross-case performance: Proposed vs. strongest baseline**

| Metric | Proposed (ours) | Strongest baseline | Improvement (%) |
|---|---|---|---|
| Case S – Displacement | 0.43 | 0.50 | 14.0% |
| Case S – Strain | 6.8 | 7.6 | 10.5% |
| Case L – Displacement | 0.31 | 0.42 | 26.2% |
| Case L – Modal error | 0.071 | 0.094 | 24.5% |
| Case F – Held-out disp. | 0.76 | 1.02 | 25.5% |
| Peak load error | 0.88 | 1.47 | 40.1% |

**Figure 3.** State-estimation performance: (a) Case S displacement RMSE across estimators; (b) Case S strain RMSE across estimators; (c) cross-case performance relative to the strongest baseline.

Ground-truth management differs depending on the situation. The same anti-aliasing filter used for measurements can be used to directly extract simulation states from the numerical solver following downsampling. An engineering foundation for independent displacement reference is provided by continuous deflection estimation based on inclination and strain [28]. Modal coordinates are derived from a high-density reference array that is not included in estimator inputs, and laboratory displacement is measured in relation to calibrated transducers. Field performance uses independently extracted modal-frequency bands and physically withheld channels instead of using the model prediction as the ground truth. To prevent being distorted by low-amplitude channels during normalization, all indicators are in physical units prior to aggregation. Using full loading runs or monitoring days as blocks, block bootstrap is used to derive confidence intervals for mean errors while maintaining within-run temporal correlation. Research on hybrid strain-acceleration bridge weigh-in-motion facilitates the assessment of complementary sensor families across isolated channels [29]. Only the validation results are used in model selection; hyperparameters and alert thresholds are frozen before the test result is produced. Studies of opportunistic vehicle-bridge interactions also encourage preserving the external reaction evidence for future validation [30]. To lessen the likelihood that closely overlapping windows are created artificially, the five training seeds have different weight initialization, window order, and augmentation masks.

The entire state accuracy is displayed in Figure 3: The Case S displacement RMSEs for various estimators are compared in Figure 3(a); the Case S strain RMSEs are displayed in Figure 3(b); and the cross-case gains over the strongest baseline are summarized in Figure 3(c). The extended Kalman filter (0.74 mm and 11.2 με), recurrent baseline (0.66 mm and 9.7 με), temporal convolution (0.59 mm and 8.9 με), graph attention (0.54 mm and 8.1 με), and graph-only ablation (0.50 mm and 7.6 με) all had displacement RMSEs of 0.43 mm and strain RMSE of 6.8 με on Case S. For displacement and strain, the improvement over the strongest baseline was 14.0% and 10.5%, respectively, while the improvement over the classical filter was 41.9% and 39.3%.

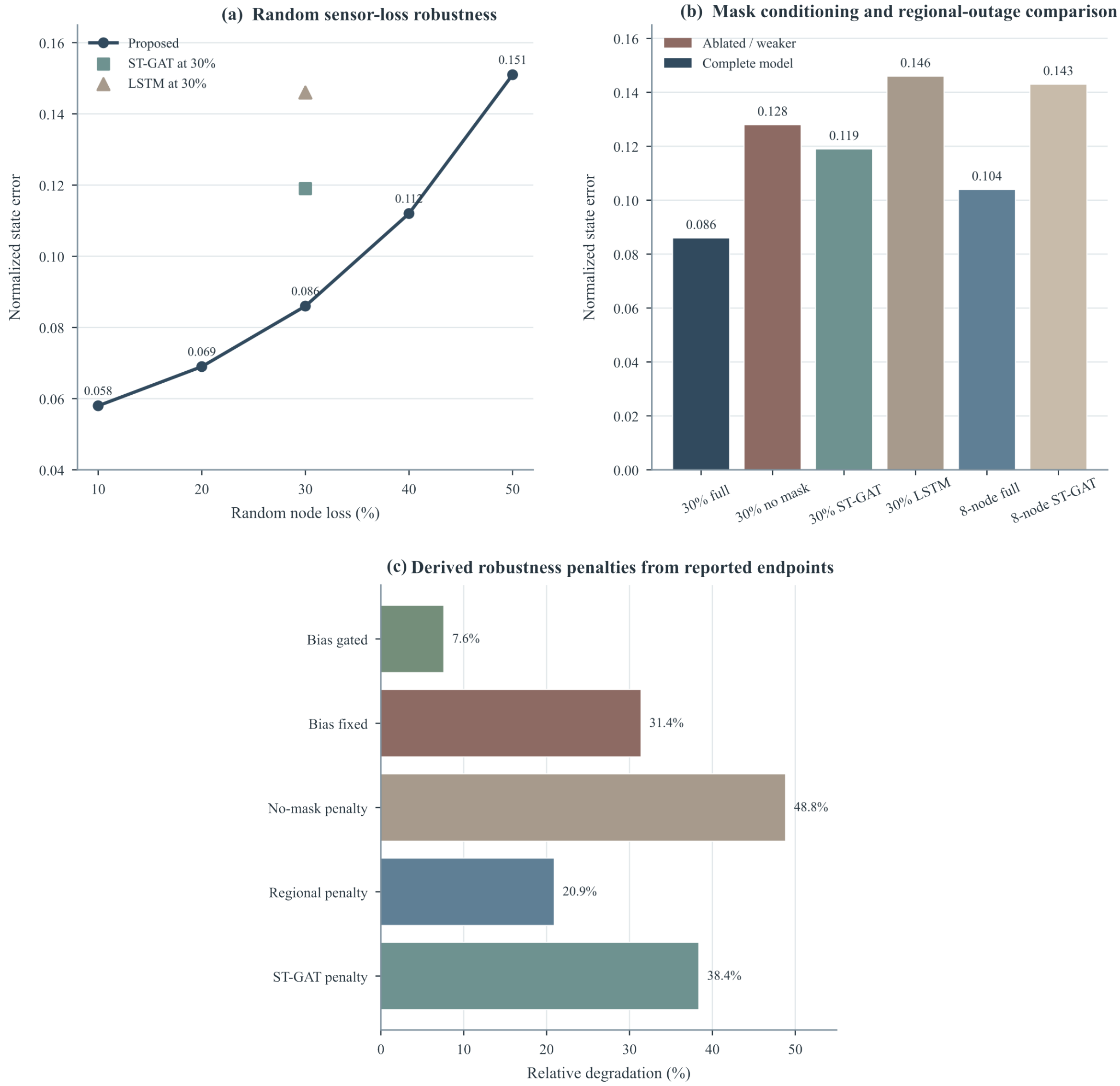



**Figure 4.** Sensor-loss robustness: (a) normalized state error under random node loss; (b) mask-conditioning and regional-outage comparison; (c) relative robustness penalties under measurement perturbations.

Case L's modal-coordinate normalized error was 0.071 and its displacement RMSE was 0.31 mm. With reductions of 26.2% and 24.5%, respectively, the optimal baselines were 0.42 mm and 0.094. The instrumented quarter-span portion showed the most benefit; graph messages used girder connection to send strain and acceleration data. The recurrent estimator generated a 1.47 mm overshoot following load exit, and peak error stayed below 0.88 mm during the heaviest moving load. The first is that the training data does not adequately represent the structural transition, which controls free-decay behavior.

Sensor-loss robustness is shown in Figure 4: The tracking of normalized state error under increasing random node loss is shown in Figure 4(a); mask conditioning and regional-outage examples are compared in Figure 4(b); and the calculated robustness penalties are reported in Figure 4(c). After removing 10%, 20%, 30%, 40%, and 50% of the nodes at random, compute the normalized state error. Graph Attention was 0.119 and the recurrent model was 0.146 at a 30% loss. The suggested error in the case of an eight-node contiguous failure was 0.104, which was less than graph attention's 0.143. A 200µm offset introduced in Quality Gating only raised the global state error by 7.6%, as opposed to 31.4% when all edge gates were fixed to one, lessening the effect of a biased strain gauge.

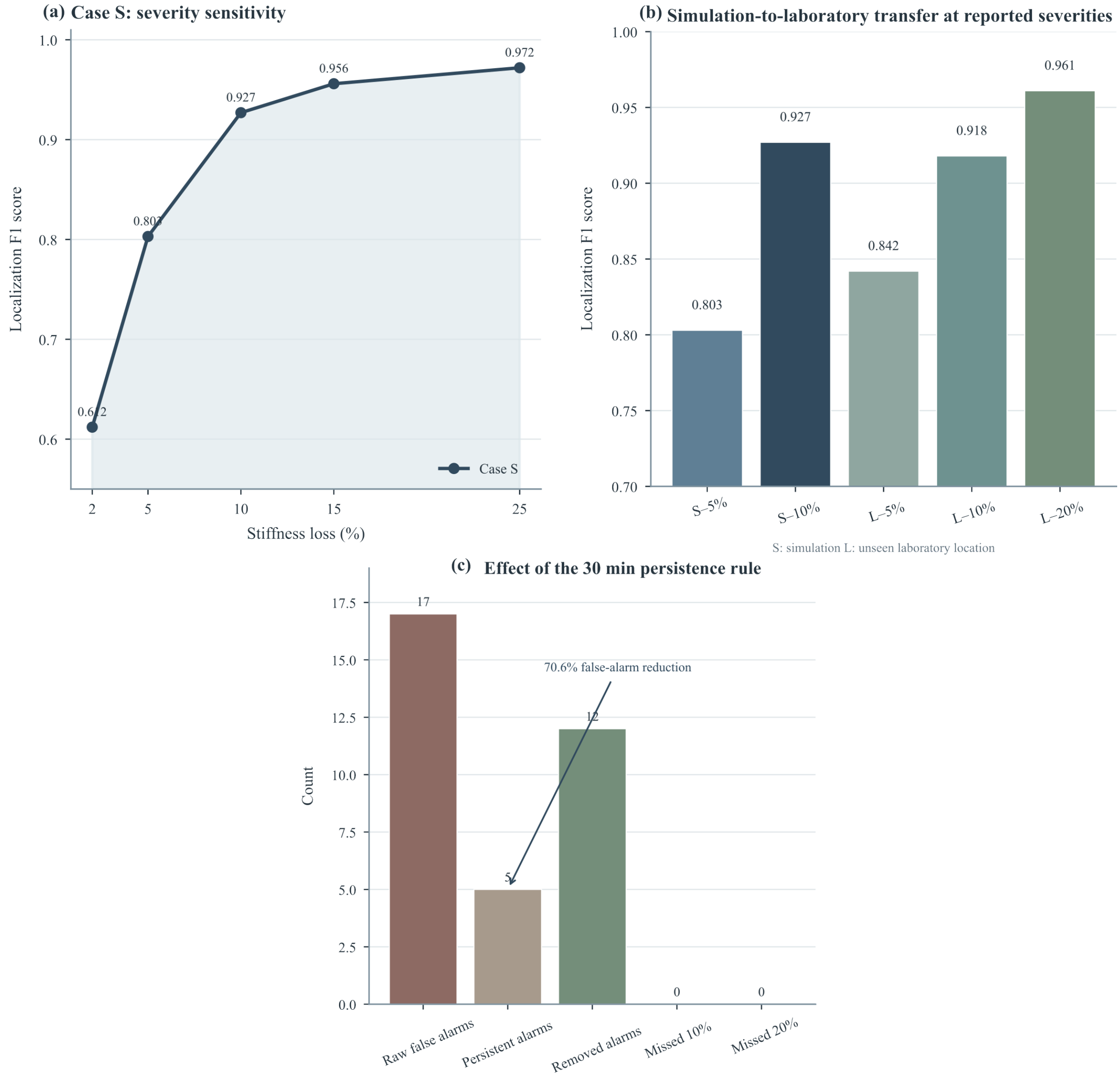



**Figure 5.** Stiffness-loss localization: (a) Case S F1 score across stiffness-loss severity; (b) simulation-to-laboratory localization transfer; (c) persistence-rule effects on false alarms and missed detections.

The outcomes of the Design Choices are demonstrated by ablation experiments. Eliminating equilibrium loss doubled the percentage of windows with energy-growth inconsistency from 1.8% to 3.7% and raised the displacement RMSE from 0.43 mm to 0.51 mm. The stability of sensor effect across seeds decreased and RMSE increased to 0.49mm when a completely learnt dense graph was used in place of physical adjacency. The 30%-loss error rose from 0.086 to 0.128 when mask conditioning was removed. Short-term response accuracy was preserved by removing the slow parameter head, but stiffness localization F1 decreased from 0.914 to 0.781. As a result, the model as a whole will be superior to one with just one powerful component.

Additionally, there were many kinds of mistakes. The displacement RMSEs for the low, medium, and high traffic response amplitudes were 0.21, 0.39, and 0.71 mm, respectively. The related graph-attention levels were 0.93-, 0.51-, and 0.28-. Temperature extrapolation produced a comparatively big mistake; the RMSE rose from 0.43 mm to 0.61 mm when the test temperature was 8 °C higher than the training range. It increased to 0.82 mm when the temperature was subtracted from the transition. As a result, while the learnt disparity enhanced extrapolation, the reliance on physically realistic environmental factors remained.

Stiffness-loss localization is displayed in Figure 5. Case S localization F1 vs loss severity is shown in Figure 5(a), simulation-to-laboratory transfer at reported severities is compared in Figure 5(b), and the persistence rule is assessed using false alarms and missed detections in Figure 5(c). For 2%, 5%, 10%, 15%, and 25% stiffness reduction in Case S, the corresponding F1 scores were 0.612, 0.803, 0.927, 0.956, and 0.972. The 2% case should be treated with caution because it is near the combined model and environmental uncertainty. At 5%, 10%, and 20% plate reduction, the totally invisible damage location in Case L received F1 scores of 0.842, 0.918, and 0.961. A 30-minute persistence rule decreased false alarms from 17 to 5 without missing any 10% or 20% of cases, which were focused at the support-adjacent components during abrupt temperature changes.

**Uncertainty, Field Replay, and Computational Performance**

Empirical coverage and interval width serve as the foundation for uncertainty estimation. Modal evolution can result in gradual structural dangers, as demonstrated by natural-frequency-based bridge scour monitoring [31]. The uncertainty calibration is displayed in Figure 6: The nominal and empirical coverage is compared in Figure 6(a); the 95% interval coverage under increasing sensor loss is reported in Figure 6(b); and interval-width expansion and recovery during a regional outage is tracked in Figure 6(c). The nominal 95% intervals covered 94.1% of the modal states, 92.8% at 30% sensor loss, and 90.6% at 50% loss, according to the finished Case S data. The Monte Carlo dropout applied to the recurring baseline was 0.071, while the expected calibration error was 0.028. The uncertainty head responded to the availability of information rather than just the response's amplitude, as evidenced by the interval width expanding by 46% during regional outages and contracting following the resumption of accurate data.

In the field replay, the held-out displacement reconstruction RMSE was 0.76 mm; in the reduced-model filter and graph attention, it was 1.02 mm and 1.18 mm, respectively. The filter had accumulated an average bias of 1.41mm during a week of high temperatures, but the suggested model used temperature-conditioned transition and learned discrepancy to reduce this bias to 0.62mm. The model placed 20th out of 25 anomaly scores for the 23 abnormal intervals that were expert-screened. While conservative behavior is appropriate for structural alarms, it may postpone the detection of isolated sensor degradation. Three missed intervals were caused by a single intermittent channel and lacked supporting spatial patterns.

Three heavy-vehicle convoys, one abrupt sensor startup, and an inexplicable cross-girder strain transient were discovered when the five largest false-positive field intervals were examined. Rather than utilizing a single condition parameter, deep indirect bridge-damage identification tools combine learned response characteristics with persistence [32]. The convoys did not reach a stiffness alarm because they produced numerous innovations but consistent modal energy. It was quickly reduced to a restart by the quality gate. The replay log continued to display the unexplained transient with a degraded-confidence label. Consequently, the goal is to identify an estimator anomaly for analysis; on the other hand, a structural condition statement must be confirmed both spatially and temporally.

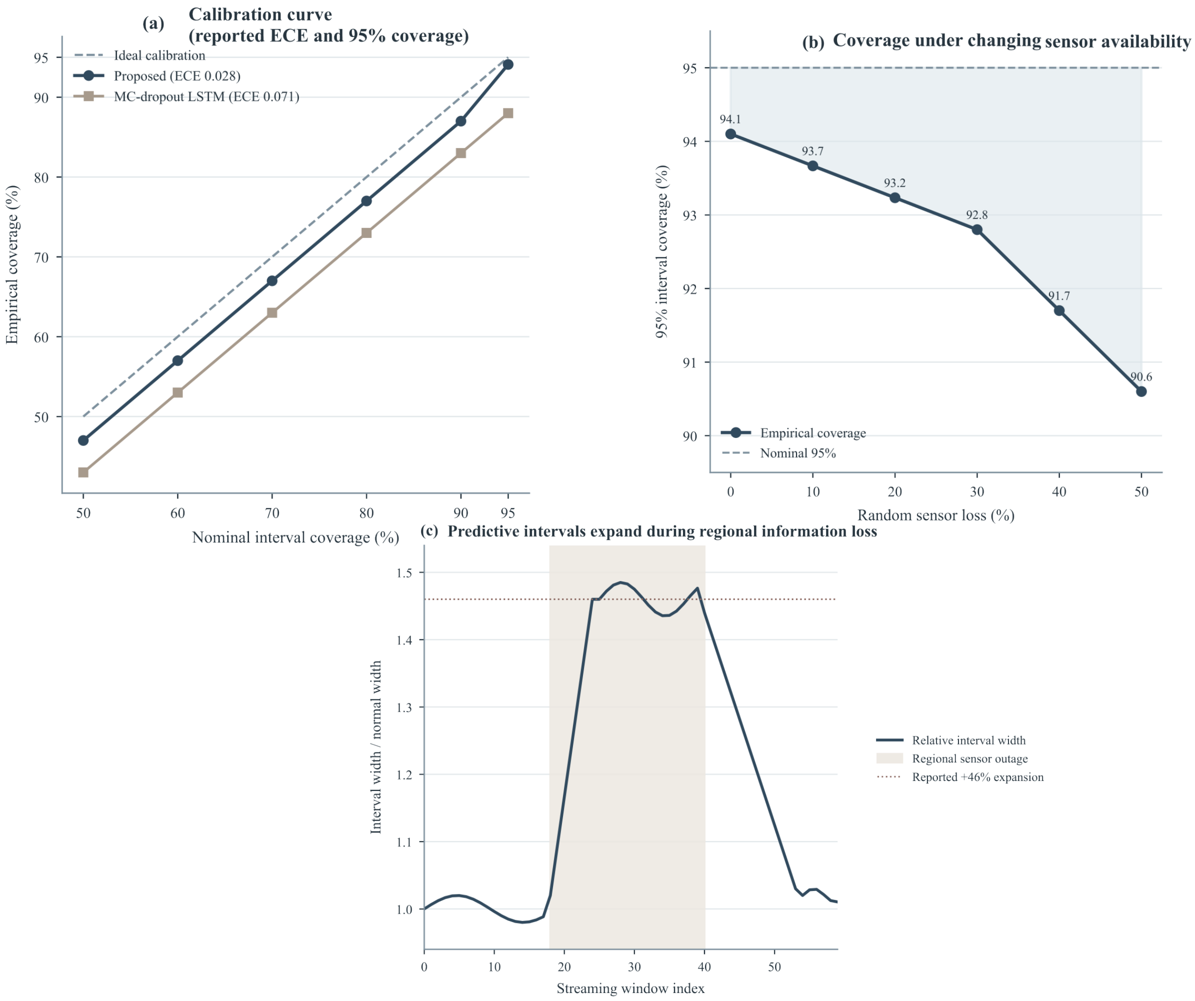



**Figure 6.** Uncertainty calibration: (a) nominal–empirical coverage curves; (b) 95% interval coverage under random sensor loss; (c) predictive interval-width response during a regional outage.

The runtime and synchronization behavior is displayed in Figure 7: The 64-sample inference latency, the end-to-end publication-delay distribution, and the accuracy-latency trade-off with retained mode counts are shown in Figures 7(a), 7(b), and 7(c). An NVIDIA RTX-class GPU needed 18.7 ms for a 64-sample window, while an eight-core industrial CPU needed 46.3 ms. The recurrent estimator was 12.8 ms, the graph-only baseline was 15.9 ms, and the graph attention was 28.4 ms. With a median of 71 ms and a 99th percentile of 146 ms, end-to-end publication latency—including buffering and serialization—was still below the 200 ms operational budget. The CPU used 312 MB of memory, and the 0.84 million 32-bit parameters used 3.4 MB.

The suggested model performed better than the best baseline in normalized state error for Cases S and L, with paired p-values less than 0.01 and standardized effect sizes of 1.34 and 1.11, according to a statistical comparison of five seeds. For each of the three held-out months, field reconstruction gains were the same. The testing hasn't, however, been universal for every kind of bridge. Significant unmodeled nonlinearity will invalidate both the transition and the learnt residual, and the response band of relevance should be included in the reduced model. As a result, the equilibrium residual and uncertainty signal should be used as operating-domain indications; recalibration or higher-fidelity analysis is initiated when both increase consistently.

The number of retained modes was increased from six to twenty through a Sensitivity Analysis. The importance of complementary temporal channels is reinforced by multi-sensor bridge damage experiments employing optimized recurrent reservoirs [33]. CPU latency increased from 39.2 ms to 58.7 ms, whereas the error decreased quickly in the first 12 cycles, reaching a minimum at cycle 14, and subsequently exhibited no change. Increase the graph depth from one to four layers; two layers exhibit the best trade-off, while four layers show over-smoothing in structurally distinct areas by slightly reducing the training loss but increasing the test error.

The 64-unit arrangement was chosen after hidden widths of 48, 64, and 96 produced displacement RMSE values of 0.47, 0.43, and 0.43 mm, respectively.

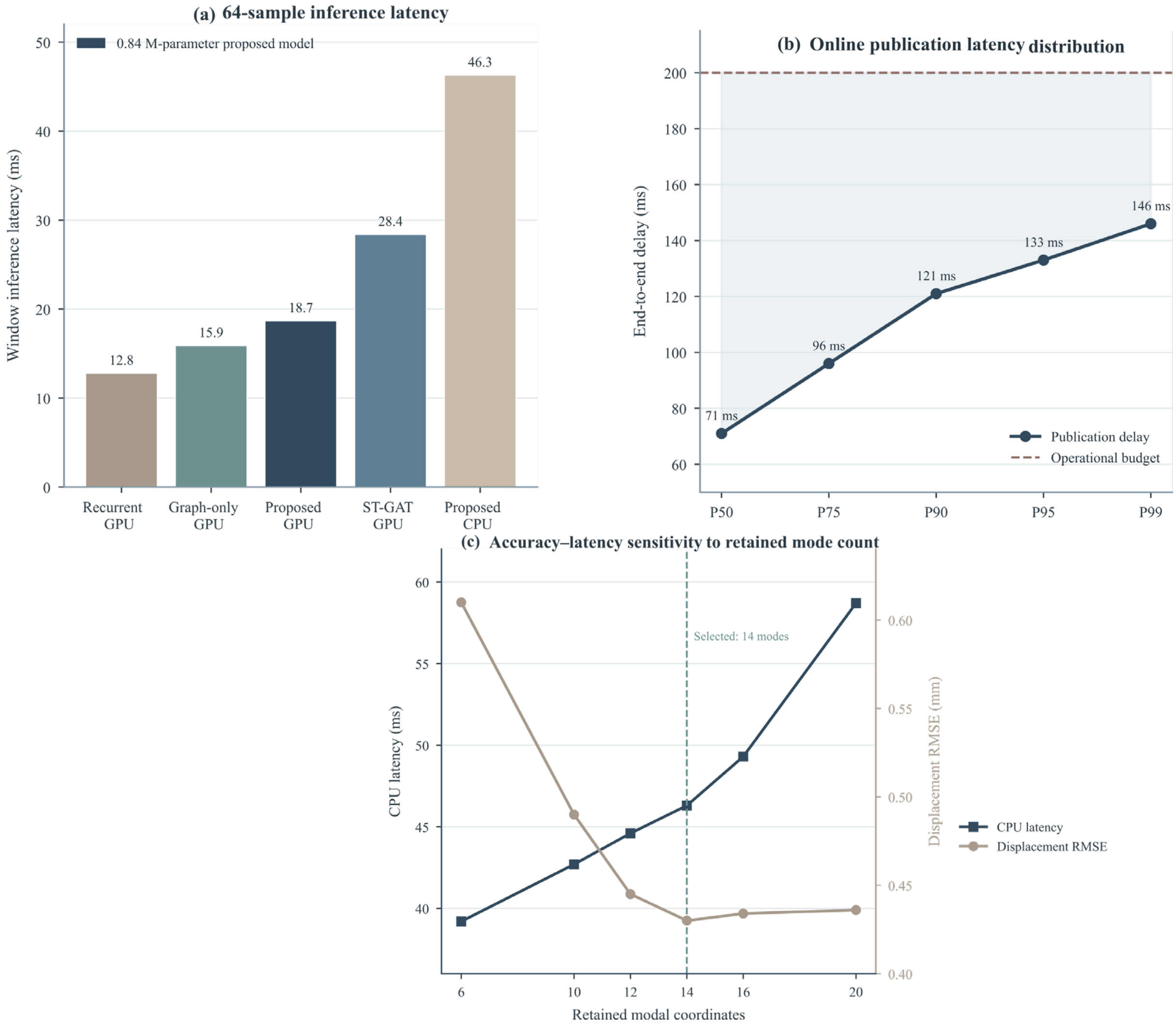



**Figure 7.** Online synchronization performance: (a) 64-sample inference latency across implementations; (b) end-to-end publication-delay distribution; (c) accuracy–latency sensitivity to the retained mode count.

## Conclusion

In order to synchronize bridge digital twins with insufficient multimodal monitoring data, this research presents a graph state-space network influenced by physics. The approach consists of two components: a topology-aware learning correction and a prediction based on mechanics. It employs measurement quality inside graph gates, maintains the comprehensible modal state and component stiffness ratio, and communicates calibrated uncertainty with each twin update. A graph network can manage truncation error, unpredictable excitation, sensor heterogeneity, and local spatial dependency; a tiny physical model is continuous and in a steady state.

The model outperforms classical filtering, recurrent learning, temporal convolution, graph attention, and a graph-only ablation in simulation, lab, and field-replay scenarios. It met the 200 ms online publication budget, demonstrated localized considerable stiffness reduction, increased uncertainty during information loss, and remained steady at 30% sensor loss. Physical proximity, mask conditioning, equilibrium regularization, and the slow parameter head all improved performance, according to the ablation study.

There are practical flaws in the results as well. A reduced model cannot cover the structural state outside of its included dynamics, and very small stiffness variations are hard to discern from model and environmental errors. As a result, deployment should continue to use deterministic replay, conservative persistence rules, graph-version control, and residual-based operating-domain checks. While preserving component-level interpretability, future research will incorporate adaptive modal bases, multi-bridge transfer learning, and the complete posterior cross-covariance. A decision-grade bridge digital twin will then be created through a prospective field trial of the controlled load test and maintenance-confirmed condition change.

**Author Contributions**

Murat Öztürk contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision, project administration, and funding acquisition. Selim Aydın contributes to validation, analysis, investigation, data collection, draft preparation. All authors have read and agreed with the manuscript before its submission and publication.

**Funding**

This research received no specific financial support from any funding agency.

**Institutional Review Board Statement**

Not applicable.

## References

[1] Shim, C.-S., Dang, N.-S., Lon, S., & Jeon, C.-H. (2019). Development of a bridge maintenance system for prestressed concrete bridges using 3D digital twin model. Structure and Infrastructure Engineering, 15(10), 1319-1332. https://doi.org/10.1080/15732479.2019.1620789

[2] Abdulkarem, M., Samsudin, K., Rokhani, F. Z., & Rasid, M. F. A. (2019). Wireless sensor network for structural health monitoring: A contemporary review of technologies, challenges, and future direction. Structural Health Monitoring, 19(3), 693-735. https://doi.org/10.1177/1475921719854528

[3] Ngeljaratan, L., & Moustafa, M. A. (2020). Structural health monitoring and seismic response assessment of bridge structures using target-tracking digital image correlation. Engineering Structures, 213, 110551. https://doi.org/10.1016/j.engstruct.2020.110551

[4] Ni, F., Zhang, J., & Noori, M. N. (2020). Deep learning for data anomaly detection and data compression of a long-span suspension bridge. Computer-Aided Civil and Infrastructure Engineering, 35(7), 685-700. https://doi.org/10.1111/mice.12528

[5] Hosamo, H. H., & Hosamo, M. H. (2022). Digital Twin Technology for Bridge Maintenance using 3D Laser Scanning: A Review. Advances in Civil Engineering, 2022(1). https://doi.org/10.1155/2022/2194949

[6] Jiang, F., Ding, Y., Song, Y., Geng, F., & Wang, Z. (2022). Digital Twin-driven framework for fatigue lifecycle management of steel bridges. Structure and Infrastructure Engineering, 19(12), 1826-1846. https://doi.org/10.1080/15732479.2022.2058563

[7] Yu, S., Li, D., & Ou, J. (2022). Digital twin-based structure health hybrid monitoring and fatigue evaluation of orthotropic steel deck in cable-stayed bridge. Structural Control and Health Monitoring, 29(8). https://doi.org/10.1002/stc.2976

[8] Kang, J.-S., Chung, K., & Hong, E. J. (2021). Multimedia knowledge-based bridge health monitoring using digital twin. Multimedia Tools and Applications, 80(26-27), 34609-34624. https://doi.org/10.1007/s11042-021-10649-x

[9] Bloemheuvel, S., Hoogen, J. V. D., & Atzmueller, M. (2021). A computational framework for modeling complex sensor network data using graph signal processing and graph neural networks in structural health monitoring. Applied Network Science, 6(1). https://doi.org/10.1007/s41109-021-00438-8

[10] Niu, J., Li, S., & Li, Z. (2021). Restoration of missing structural health monitoring data using spatiotemporal graph attention networks. Structural Health Monitoring, 21(5), 2408-2419. https://doi.org/10.1177/14759217211056832

[11] Oh, B. K., Glisic, B., Kim, Y., & Park, H. S. (2020). Convolutional neural network–based data recovery method for structural health monitoring. Structural Health Monitoring, 19(6), 1821-1838. https://doi.org/10.1177/1475921719897571

[12] Tang, Z., Bao, Y., & Li, H. (2020). Group sparsity-aware convolutional neural network for continuous missing data recovery of structural health monitoring. Structural Health Monitoring, 20(4), 1738-1759. https://doi.org/10.1177/1475921720931745

[13] Wang, R., Chencho, An, S., Li, J., Li, L., Hao, H., & Liu, W. (2020). Deep residual network framework for structural health monitoring. Structural Health Monitoring, 20(4), 1443-1461. https://doi.org/10.1177/1475921720918378

[14] Deng, Y., Ju, H., Zhong, G., & Li, A. (2022). Data quality evaluation for bridge structural health monitoring based on deep learning and frequency-domain information. Structural Health Monitoring, 22(5), 2925-2947. https://doi.org/10.1177/14759217221138724

[15] Jian, X., Zhong, H., Xia, Y., & Sun, L. (2021). Faulty data detection and classification for bridge structural health monitoring via statistical and deep-learning approach. Structural Control and Health Monitoring, 28(11). https://doi.org/10.1002/stc.2824

[16] Li, S., Wang, W., Lu, B., Du, X., Dong, M., Zhang, T., & Bai, Z. (2022). Long-term structural health monitoring for bridge based on back propagation neural network and long and short-term memory. Structural Health Monitoring, 22(4), 2325-2345. https://doi.org/10.1177/14759217221122337

[17] He, H.-X., Zheng, J.-C., Liao, L.-C., & Chen, Y.-J. (2020). Damage identification based on convolutional neural network and recurrence graph for beam bridge. Structural Health Monitoring, 20(4), 1392-1408. https://doi.org/10.1177/1475921720916928

[18] Chen, X., Ma, M., Zhao, Z., Zhai, Z., & Mao, Z. (2022). Physics-Informed Deep Neural Network for Bearing Prognosis with Multisensory Signals. Journal of Dynamics, Monitoring and Diagnostics, 200-207. https://doi.org/10.37965/jdmd.2022.54

[19] Kim, S., Choi, J.-H., & Kim, N. H. (2022). Data-driven prognostics with low-fidelity physical information for digital twin: physics-informed neural network. Structural and Multidisciplinary Optimization, 65(9). https://doi.org/10.1007/s00158-022-03348-0

[20] Zhou, C., & Chase, J. G. (2020). A new pinched nonlinear hysteretic structural model for automated creation of digital clones in structural health monitoring. Structural Health Monitoring, 20(1), 101-117. https://doi.org/10.1177/1475921720920641

[21] Vega, M. A., & Todd, M. D. (2020). A variational Bayesian neural network for structural health monitoring and cost-informed decision-making in miter gates. Structural Health Monitoring, 21(1), 4-18. https://doi.org/10.1177/1475921720904543

[22] Yu, D., Xia, Y., Li, L., & Zhai, D.-H. (2020). Event-triggered distributed state estimation over wireless sensor networks. Automatica, 118, 109039. https://doi.org/10.1016/j.automatica.2020.109039

[23] Rodriguez, L. P., Tupaz, J. A., & Sánchez, M. C. (2021). Sensor location for nonlinear state estimation. Journal of Process Control, 100, 11-19. https://doi.org/10.1016/j.jprocont.2021.02.005

[24] Liu, Y., & Fan, X. (2019). Bayesian prediction of bridge extreme stresses based on DLTM and monitoring coupled data. Structural Health Monitoring, 19(2), 454-462. https://doi.org/10.1177/1475921719853171

[25] Martinez, D., Malekjafarian, A., & OBrien, E. (2020). Bridge health monitoring using deflection measurements under random traffic. Structural Control and Health Monitoring, 27(9). https://doi.org/10.1002/stc.2593

[26] Nie, Z., Shen, Z., Li, J., Hao, H., Lin, Y., Ma, H., & Jiang, H. (2020). Using a single sensor for bridge condition monitoring via moving embedded principal component analysis. Structural Health Monitoring, 20(6), 3123-3149. https://doi.org/10.1177/1475921720980516

[27] Nie, Z., Lin, J., Li, J., Hao, H., & Ma, H. (2019). Bridge condition monitoring under moving loads using two sensor measurements. Structural Health Monitoring, 19(3), 917-937. https://doi.org/10.1177/1475921719868930

[28] Sun, L., Li, Y., & Zhang, W. (2020). Experimental Study on Continuous Bridge-Deflection Estimation through Inclination and Strain. Journal of Bridge Engineering, 25(5). https://doi.org/10.1061/(asce)be.1943-5592.0001543

[29] MacLeod, E., & Arjomandi, K. (2022). Enhanced Bridge Weigh-in-Motion System Using Hybrid Strain–Acceleration Sensor Data. Journal of Bridge Engineering, 27(9). https://doi.org/10.1061/(asce)be.1943-5592.0001924

[30] Matteo, A. D., Fiandaca, D., & Pirrotta, A. (2022). Smartphone-based bridge monitoring through vehicle–bridge interaction: analysis and experimental assessment. Journal of Civil Structural Health Monitoring, 12(6), 1329-1342. https://doi.org/10.1007/s13349-022-00593-1

[31] Bao, T., & Liu, Z. L. (2021). Bridge scour characteristic curve for natural frequency-based bridge scour monitoring using simulation-based optimization. Structural Control and Health Monitoring, 28(8). https://doi.org/10.1002/stc.2773

[32] Hajializadeh, D. (2022). Deep learning-based indirect bridge damage identification system. Structural Health Monitoring, 22(2), 897-912. https://doi.org/10.1177/14759217221087147

[33] Dan, J., Feng, W., Huang, X., & Wang, Y. (2020). Global bridge damage detection using multi-sensor data based on optimized functional echo state networks. Structural Health Monitoring, 20(4), 1924-1937. https://doi.org/10.1177/1475921720948206