Journal of Applied Computational Intelligence

eISSN: 3136-6937

Vol. 4, No. 2, April, 2026

Bridge Weigh-in-Motion Axle Load Identification Using a Graph Attention Network with Moving-Load Influence-Line Encoding

Klemen Majcen[1], Alenka Kobal[2], Simona Pečnik[2,*]

[1] *Faculty of Civil Engineering, Transportation Engineering and Architecture, University of Maribor, Maribor, 2000, Slovenia*

[2] *Department of Artificial Intelligence and Intelligent Systems, Faculty of Computer and Information Science, University of Ljubljana, Ljubljana, 1000, Slovenia*

Received 08 December 2025; Revised 30 March 2026; Accepted 08 April 2026; Published 12 April 2026

Abstract

Bridge weigh in motion requires individual axle loads to be separated from structural responses that overlap in time and vary with vehicle position. This paper develops an influence line encoded graph attention framework that represents sensor windows as graph nodes and conditions message exchange on moving axle coordinates. A position dependent encoder supplies influence ordinates, local slopes, and axle separation features, while a positive regression head estimates static axle loads. A response consistency term links the predicted axle loads to reconstruction of the measured strain envelope. The evaluation benchmark comprises 12000 vehicle passages, six bridge strain channels, five speed levels, and three noise conditions. The comparison gives the proposed framework a mean axle absolute percentage error of 3.20%, compared with 6.10% for a graph attention model without influence encoding and 8.70% for regularized inversion. The 95th percentile axle error is 8.10%, and the corresponding vehicle gross weight error is 1.80%. Controlled sensitivity scenarios raise mean axle error to 5.80% under a 15% influence line perturbation and to 4.60% when two channels are unavailable. The evaluation examines overall accuracy together with condition dependent degradation and the distribution of axle errors. The framework provides a physically interpretable route for combining bridge response geometry with adaptive relational learning, and defines the calibration, uncertainty, and validation checks required before field deployment. The reported errors describe the specified benchmark and do not establish a certified weighing class or an operational acceptance threshold.

Keywords: Bridge Weigh in Motion, Axle Load Identification, Graph Attention Network, Influence Line Encoding, Moving Load Inversion

1. Introduction

Weigh-in-motion research distinguishes axle-level estimation from gross vehicle weight estimation and examines the effects of vehicle dynamics on weighing accuracy [1]. Influence-line-based identification relates moving axle loads to the measured bridge response through a position-dependent response operator [2]. Vehicle–bridge interaction models show that pavement properties, road roughness, vehicle speed, and structural parameters influence the coupled dynamic response [3]. Closely spaced axles pose a persistent challenge for bridge weighing, motivating sequential neural models that combine vehicle-level and axle-level information [4]. This distinction matters for engineering use: a small error in gross weight can coexist with substantial redistribution error between adjacent axle estimates. A useful identification

* Corresponding author: pecink.si@um.si

http://dx.doi.org/10.68406/jaci.2026.vol4iss2no2:14-29

➢ This is an open access article under the CC BY-NC-SA 4.0 license (https://creativecommons.org/licenses/by/4.0/).

© Authors retain all copyrights.

framework must therefore retain the mechanics of load transfer while resolving the temporal and spatial ambiguity created by overlapping axle responses. It must also identify conditions in which limited observational information weakens axle separation. Such requirements favor models whose inputs include the changing physical relationship between loads and sensors, and explicitly represent the spatial and temporal structure of each passage. The relationship between axle spacing and sensor sensitivity is especially important when several axles occupy the sensitive region simultaneously. Assessment therefore needs to retain the signed force allocation error for each axle as well as the total vehicle error. They further require evaluation at the axle, vehicle, and operating condition levels, because a single aggregate score can conceal errors that are relevant to bridge assessment.

Conventional inverse formulations compare a measured response with a superposition of shifted influence signatures and infer the axle forces that best reproduce it. Load-identification methods have combined wavelet denoising with regularization to reduce noise interference and improve the conditioning of inverse force recovery [5]. The chosen penalty also shapes how an ambiguous vehicle load is redistributed, particularly when tandem axles produce similar signatures over most of the crossing. Data driven regressors offer a different route by learning nonlinear associations between response segments and target loads. Hussain et al. [6] used deep convolutional networks to identify vehicle loads from bridge responses while accounting for variations in speed, road roughness, and measurement noise. Their practical attraction is the ability to reuse patterns across many passages, but a learned association can reflect the load and speed frequencies of the training set at the expense of transferable force–response relationships. Position information is especially vulnerable to this problem. The same axle can produce different strain amplitudes at different coordinates, whereas different axles can produce similar amplitudes at different times. Representing this changing relationship requires information beyond a fixed sensor ordering. If position is supplied only as an auxiliary scalar, the model must discover both the mechanical influence pattern and the load mapping from the same finite set of examples. The resulting burden is greatest for sparsely represented vehicles and operating conditions outside the calibration range. An engineering model should reserve learning capacity for residual relationships that remain after accessible mechanical information has been represented.

Graph learning provides a way to organize relational information without requiring every sensor pair to contribute equally. Cross-attention-based graph networks have been developed to fuse multi-sensor vibration features for fault diagnosis [7]. Interpreting the path from an axle to measured strain requires mechanically defined features alongside attention coefficients. Physics-informed neural networks integrate physical laws into learning objectives for forward and inverse engineering problems [8]. For bridge weighing, the relevant context changes throughout a passage: an axle enters the influence region, approaches a sensitive section, and eventually leaves the observed span. Neighboring windows should exchange information when they describe compatible portions of that trajectory, using temporal proximity together with physical sensitivity. The present study formulates this requirement as moving load influence line encoding within a graph attention network. Its intended contribution is a structured axle load estimator in which calibrated influence ordinates and slopes condition relational weighting, sensor reliability is represented through masks, and a response consistency penalty links predicted forces back to the observation domain. The formulation separates a physically anchored response path from a learned correction, with perturbation and masking scenarios assessing the approximation under changing dynamic conditions. The methodological contribution is the use of axle-position-dependent influence features to condition graph communication, coupled with a separate response-consistency objective.

The following sections develop the proposed framework from the mechanics of moving load observation to the assessment of axle identification accuracy, with particular attention to the effects of response overlap, influence calibration, and incomplete sensing. The evaluation design links independently calibrated bridge responses, reference axle measurements, and condition specific error metrics. This sequence connects the representation of structural sensitivity with the evaluation of individual axle errors, providing a consistent basis for interpreting the proposed learning components. The scientific content is organized into six sections, followed by declarations and references. Section 2 reviews influence line inversion, learned response representations, and relational modeling, identifying the specific gap addressed by the proposed encoder. Section 3 defines the observation model, moving coordinate system, and graph construction, including the separation between static sensitivity and dynamic residuals. Section 4 develops the influence conditioned attention mechanism, positive axle regression, training objective, and inference procedure. Section 5 specifies an evaluation protocol and examines the benchmark through accuracy distributions, vehicle strata, component ablation, robustness, and computational reporting. The evaluation links each metric to a defined operating condition and identifies the independent measurements required to test the corresponding interpretation. Section 6 presents the methodological conclusions, acknowledges the unresolved limitations, and identifies the field investigations needed to assess the framework under realistic bridge and traffic conditions.

2. Related Work

2.1. Influence line inversion and axle separation

Influence line methods express bridge response as a load weighted combination of position dependent sensitivities. Their value lies in the direct connection between the unknown axle forces and a calibrated structural response pattern. Influence-line identification has been formulated jointly with vehicle-load estimation by embedding axle-load–response equations in physics-informed neural networks [9]. A well calibrated response operator supports transparent axle recovery, while the distinction between neighboring axle signatures determines the available separation information. Closely spaced axles may generate strongly correlated contributions even when the forward response is accurately represented. This issue concerns both information content and numerical optimization. Adding a complementary measurement section can supply more independent information than densely sampling nearly identical response shapes. The calibration procedure must therefore be considered jointly with the intended vehicle population and sensor arrangement. A method that appears stable for isolated axles may remain poorly conditioned for compact axle groups, particularly when the observation duration is shortened or the response is filtered aggressively.

Iterative regularization has been combined with a Newmark-β response formulation to stabilize dynamic load identification under measurement noise [10]. Such regularization can reduce noise amplification, although its influence on individual axle allocation deserves separate examination. A penalty that yields a plausible gross weight can still favor an inaccurate split within a tandem group. Dynamic compensation introduces an additional modeling decision because the oscillatory and quasi-static components generally have different relationships to static axle force. Outside bridge weighing, D-optimal strain-sensor placement has been studied for inverse load estimation on wind turbine blades, with the objective of limiting uncertainty propagation [11]. For the present framework, this literature motivates explicit response reconstruction during learning and a soft consistency constraint that accommodates dynamic residuals. The proposed encoder uses the influence line as a structured source of sensitivity information, with an explicit perturbation protocol to examine dependence on its calibration. The representation accommodates residual dynamics and evaluates axle recovery separately from response fit.

2.2. Learned temporal and spatial response representations

Learned axle identification systems can combine information across a passage without requiring a fixed analytical correction for every disturbance. Their representations may exploit local shape, phase relationships, sensor amplitude ratios, and vehicle timing. Temporal modeling in bridge monitoring has combined bidirectional gated recurrent units with attention to represent dependencies in multichannel observations [12]. These approaches require careful alignment of the learning task with the engineering target. Static axle load, instantaneous wheel force, and vehicle gross weight require distinct labels to preserve the physical interpretation of each output. Training data must also represent how observations are acquired. A representation learned from perfectly synchronized numerical channels may behave differently when sensors have timing offsets or when axle entry times are uncertain. Controlled alignment, equivalent input information, and independent passage partitions are necessary to attribute performance differences to the architecture.

Spatial learning introduces relationships among measurement channels, yet the meaning of distance depends on the observation mechanism. Sensors close together along a girder may provide highly redundant signals, whereas a more distant sensor can supply useful contrast because its influence line differs substantially. Masciotta et al. [13] combined physics-based models with data-driven approaches to optimize sensor placement for bridge structural health monitoring. This motivates a representation that contains both geometric location and response sensitivity rather than one that encodes coordinates alone. A purely temporal network can receive the same information through a sufficiently rich input, but it must learn which cross-channel comparisons remain meaningful as the vehicle moves. Conversely, a fixed spatial graph can preserve the sensor layout while missing the axle trajectory that causes changing response relationships. A comparison between these alternatives should therefore control the input information and training budget. Equivalent physical inputs are essential for attributing a performance difference to graph learning. The proposed evaluation includes a graph model with identical basic signal inputs but no influence encoding, and a separate feature ablation that removes the slope channel while preserving the remaining architecture. Sensor-placement and monitoring studies motivate the information representation; direct evidence of axle-load identification performance comes from task-matched evaluation.

2.3. Graph attention and mechanically informed encoding

Graph attention models can determine inter-node attention from learned feature covariance, as demonstrated by a Mahalanobis-distance formulation [14]. This mechanism is suitable for heterogeneous observation quality because contributions can vary across sensors and windows. Nevertheless, the learned weights depend on the graph definition and on the information available to the attention function. A graph constructed entirely from temporal proximity may propagate correlated disturbance as readily as useful load information. Dense connectivity can further obscure whether

the network is using a meaningful structural relationship or compensating for an inadequate node representation. Interpretation requires controlled feature removal, stability checks, and comparison with physically meaningful sensitivity patterns. When channels provide interchangeable information, sensor importance should be interpreted at the level of complementary channel groups.

In structural dynamics, physics-informed networks have been used to learn interpretable response models from sparse shake-table records [15]. The design choice is consequential: a hard physical constraint can impose model bias when the forward operator is approximate, while an unconstrained feature may be ignored if a spurious statistical shortcut is easier to learn. The present formulation uses influence information in two distinct roles. Local ordinates and slopes condition message construction, and a soft reconstruction term tests whether predicted loads remain compatible with the quasi-static response envelope. The graph follows moving load coordinates rather than relying solely on stationary sensor adjacency. This arrangement defines a specific methodological gap between ordinary graph weighting and explicit influence line inversion. The unresolved question is whether physical conditioning improves axle separation under distribution shifts without increasing sensitivity to calibration error. An adequate investigation must examine this tradeoff through matched baselines, blocked data partitions, perturbations of the encoder, and distributions of individual axle errors. Addressing this engineering question requires condition specific comparisons alongside the aggregate error. Feature conditioning supplies local sensitivity information before aggregation, whereas response consistency constrains the compatibility of the final axle estimates with the observed envelope.

3. Moving Load Representation and Graph Construction

3.1. Observation model and moving coordinates

The target variable is the static vertical load carried by each axle during a single vehicle passage. Axle count, entry time, longitudinal spacing, and speed trajectory are externally estimated inputs whose uncertainties are retained in the evaluation. The bridge is represented by six longitudinal strain channels at distinct observation sections. Each channel has a calibrated quasi-static influence line, expressed in strain per unit force. A sign convention is fixed before model fitting so that compression and tension are interpreted consistently across sensors. The measured strain contains a force related component, a dynamic residual, baseline drift, and measurement noise. This approximate decomposition separates the nominal load contribution from unresolved response components. The observation equation expresses the measured response as the estimated quantity is a passage-level static axle load; time-varying contact-force fluctuations contribute to the unresolved dynamic response.

$$y_s(t) = \sum_a P_a I_s(x_a(t)) + d_s(t) + \varepsilon_s(t) \tag{1}$$

Here the channel index s selects a sensor, the axle index a identifies an axle, P denotes static force, I denotes the corresponding influence ordinate, d denotes unresolved dynamic response, and ε denotes acquisition error. The moving coordinate x is computed by integrating the estimated speed from the axle entry time and subtracting its longitudinal offset within the vehicle. Outside the calibrated influence region, the ordinate is set to zero using a short continuous taper rather than an abrupt discontinuity. Variable speed is retained in this coordinate mapping; replacing it by a passage average is an explicit ablation option. Response channels are normalized with training partition scale factors, which remain fixed for validation and testing.

Passage segmentation begins before the leading axle enters the influence region and ends after the trailing axle exits. The retained margins support baseline estimation and separation of residual oscillation from the load-related response. Synchronization uses a common acquisition clock in the reference configuration, while perturbed timing is assessed separately. A channel is rejected when clipping, persistent dropout, or an implausible baseline shift exceeds a predetermined acquisition rule. Explicit validity masks distinguish missing observations from measured zero strain. The vehicle record stores sensor positions, influence calibration identifiers, timing confidence, and the mask history. Filtering is identical across learned baselines and the proposed model, with parameters selected using the training and validation partitions.

3.2. Influence encoding and feature normalization

Each node represents a sensor and a short time window. Signal features include the local mean, root mean square, signed peak, and a compact temporal embedding computed by a shared one-dimensional convolution. The physical encoder evaluates each candidate axle trajectory at the window center. It provides the influence ordinate, a spatial derivative estimated from a smoothed calibration curve, a normalized moving coordinate, and an indicator of whether the axle lies within the sensitive region. Axle spacing is retained as a vehicle level descriptor and used to form relative trajectory features. Influence ordinates, local slopes, moving coordinates, and validity indicators are assembled through the feature encoding equation

$$\mathbf{e}_{sa}(t) = \left[I_s(x_a(t)), \ell I'_s(x_a(t)), \frac{x_a(t)}{L}, m_{sa}(t) \right] \quad (2)$$

The characteristic length ℓ makes the derivative channel comparable in scale to the ordinate channel, L denotes span length, and m is the validity mask. Both length scales are specified from the bridge geometry rather than optimized against test errors. Ordinates and derivatives are scaled using calibration statistics from the training bridge. When the influence line changes sign, the signed feature is retained because replacing it by an absolute value would discard information relevant to response superposition. Derivative estimates are computed after smoothing with a fixed spatial bandwidth, avoiding amplification of small calibration irregularities. The encoder can process an arbitrary axle count by pooling axle specific features with a permutation consistent operation while preserving ordered axle queries for the final output. Padding masks distinguish absent axles from physically present axles with weak local sensitivity. A signed local slope helps distinguish rising and falling portions of an influence profile that share similar ordinates, provided that calibration resolution supports stable differentiation.

The acquisition and graph preparation workflow are presented in Figure 1, connecting strain windows, trajectory estimates, influence evaluation, and masked node assembly before model inference. The workflow distinguishes influence line calibration from attention parameter learning. A calibration file is a versioned input to the pipeline, and a new file triggers an explicit compatibility check before inference. Sensitivity thresholds are defined relative to the calibrated channel scale to control amplification in low-amplitude windows. Coordinates may be known more accurately than the influence ordinates, or the reverse; uncertainty in these inputs is therefore perturbed independently in the evaluation plan. Shift and scale perturbations are examined separately because they create different axle allocation errors. Compatibility checking should compare sensor identity, strain sign convention, spatial units, and preprocessing settings with those expected by the trained model.

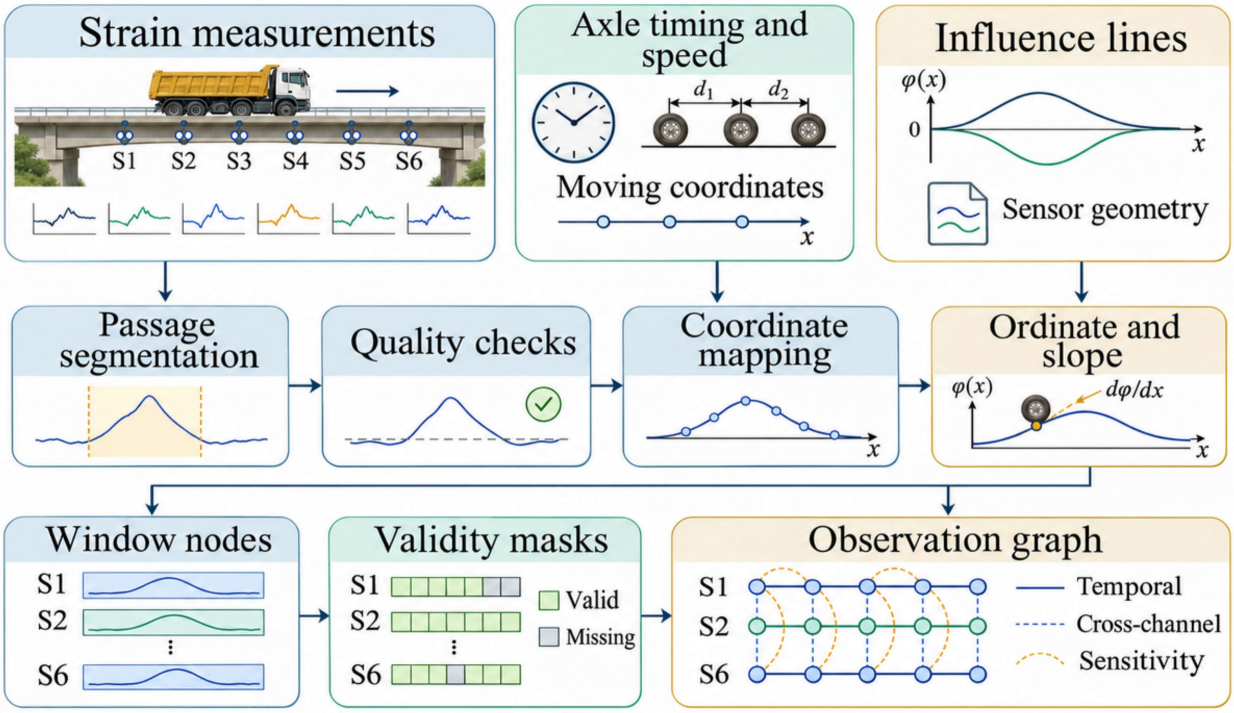


Figure 1. Moving load observation and influence encoding workflow

3.3. Sparse graph topology and information boundaries

Edges connect adjacent windows within a sensor channel, simultaneous windows across complementary channels, and windows whose moving coordinate features indicate overlapping axle sensitivity. Self-loops preserve the local observation. A sparse topology gives each node a bounded temporal neighborhood and a limited number of cross-channel connections. A candidate cross-channel edge is admitted only when both observations are valid and their influence regions overlap for at least one axle. Neighbor selection uses observable mechanical features, preserving the separation between graph construction and target load supervision. Relative time, relative longitudinal position, and influence similarity are stored as edge attributes. Edge admission depends on estimated trajectories, calibration features, and validity masks; reference axle loads are excluded from the edge-selection rule.

Graph construction is performed independently for each passage, with dataset partitions represented by disconnected graphs. The training pipeline batches passage graphs as disconnected components, preserving their information

boundaries. This boundary also applies to normalization statistics, calibration fitting, augmentation parameters, and hyperparameter selection. The graph is built from information available at deployment, whereas target axle forces enter only the supervised loss. Missing sensors remove or mask associated edges while preserving the ordering of the output axle queries. When few informative channels remain, the model returns a quality flag indicating reduced observational support. The flag describes observation quality, while prediction uncertainty requires separate calibration. An operational system would require a separate acceptance policy based on observed validation performance under the corresponding channel configuration. A quality flag records insufficient input support; a calibrated uncertainty interval additionally requires validation of coverage for the relevant operating group.

4. Influence Conditioned Graph Attention Model

4.1. Message weighting and positive axle regression

The model concatenates the temporal embedding, influence encoding, sensor identity embedding, and quality mask at each node. A shared linear transformation maps this concatenation to a latent space, followed by two residual attention blocks. Each block uses four heads with a latent width of 32 per head in the specified reference implementation. Edge attributes enter the attention score through a separate transformation, allowing moving coordinate compatibility to modify the contribution of a neighboring observation. For each receiving node, the following attention equation normalizes the contributions of its valid neighbors:

$$\alpha_{ij} = \mathrm{softmax}_j\{\mathrm{LeakyReLU}^{\top}[\mathbf{a}^{\top}(\mathbf{W}\mathbf{h}_i \| \mathbf{W}\mathbf{h}_j \| \mathbf{U}\mathbf{r}_{ij})]\} \tag{3}$$

The feature **h** denotes the node state, **r** denotes edge attributes, **W** and **U** are learned transformations, and the concatenation symbol joins compatible latent vectors. Normalization is performed only across valid neighbors of a receiving node. Head outputs are concatenated in the first block and averaged in the second, after which residual projection and normalization stabilize the feature scale. Shuffling or removing edge attributes tests whether the model uses the encoded positional relationships. Attention coefficients are retained as diagnostics for comparison with sensitivity and acquisition quality. Attention magnitude measures the allocation of learned message weights and requires comparison with feature-removal tests before it can support a mechanical interpretation.

Axle queries are constructed from the estimated entry time and spacing of each axle. They pool node states using a masked compatibility score, producing one latent representation per physical axle. The same regression head is applied to every axle query, which reduces dependence on a fixed vehicle class. A smooth positive transformation converts the scalar output to a force estimate, with a small numerical floor specified in the implementation. The architecture and training branches appear in Figure 2, including the temporal encoder, influence conditioned attention blocks, axle queries, positive load head, and response reconstruction branch. Positive output enforces physical admissibility; axle separation accuracy requires additional evaluation. If two axle queries receive nearly identical evidence, the model may redistribute loads incorrectly between axles despite an accurate gross weight estimate.

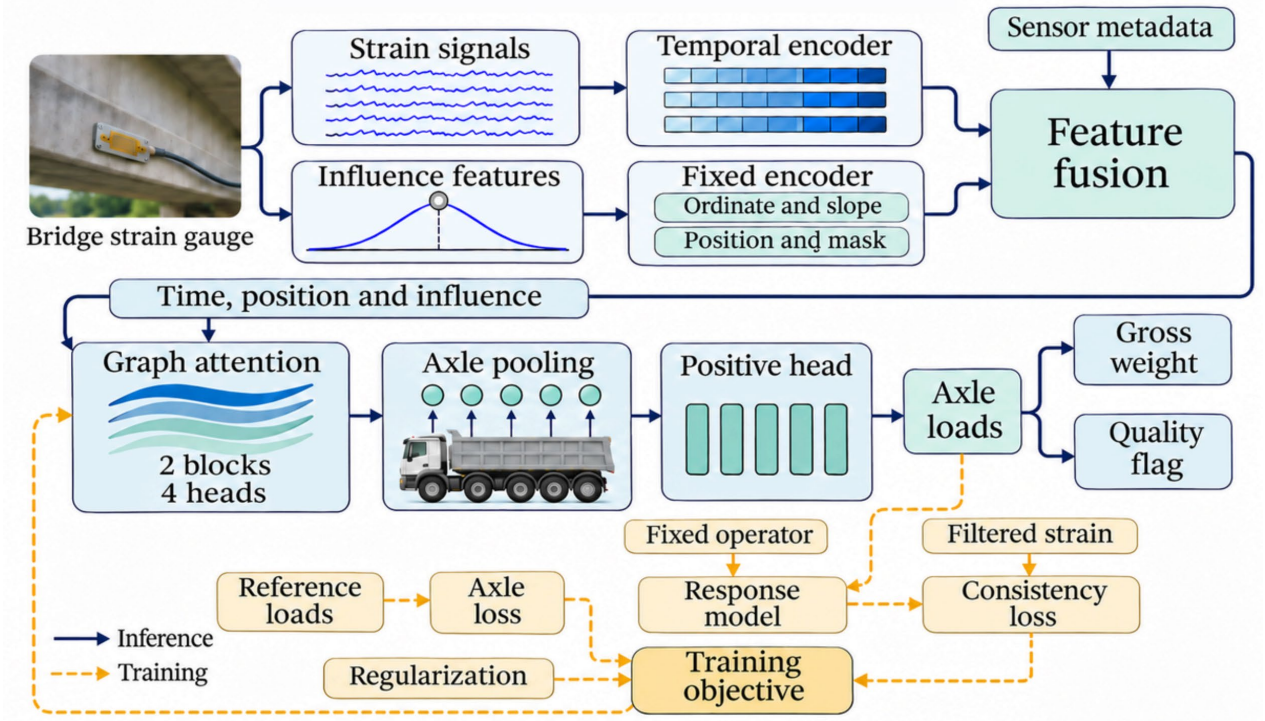


Figure 2. Influence conditioned graph attention architecture and training workflow

4.2. Objective design and training protocol

The training objective combines axle supervision, quasi-static response consistency, and parameter regularization. Relative axle error is stabilized by a fixed denominator floor to avoid excessive weighting of unusually light axles. The reconstruction target is a filtered strain envelope, reflecting the quasi-static contribution represented by the influence operator while limiting the contribution of unresolved vehicle–bridge dynamics. These terms enter a joint objective, whose loss equation is written as

$$\mathcal{L} = \mathrm{mean}_{a}\rho\left(\frac{\widehat{P}_a - P_a}{P_a^*}\right) + \lambda \mathrm{mean}_{s,t}\rho\left(\frac{\hat{y}_{st} - \bar{y}_{st}}{\sigma_s}\right) + \mu\|\boldsymbol{\theta}\|_2^2 \quad (4)$$

Here ρ is the Huber penalty, the starred force is the larger of the target force and a fixed force floor, the overbar indicates the filtered target, and σ is a channel scale determined from the training partition. The reconstructed response is obtained by superposing nominal influence signatures using the predicted static loads. The regularization coefficient μ and consistency weight λ are selected using validation passages alone. The reference design sets the consistency weight to 0.15 and the parameter penalty to 0.0001; validation selects the values appropriate to each calibration condition. A zero-consistency weight is retained as an ablation. Excessive consistency weight would encourage the model to explain dynamic residuals through static forces, which is precisely the mismatch the filtered target is intended to reduce. Channel scaling renders reconstruction errors comparable across sensors, while the consistency weight controls their contribution relative to normalized axle supervision.

Training is specified for at most 150 epochs with an adaptive optimizer, an initial learning rate of 0.001, and early stopping after 15 epochs without validation improvement. The stopping criterion uses mean vehicle-balanced axle error to assess load identification accuracy independently of the reconstruction term. A five-seed training protocol supports analysis of optimization variability, with vehicle-level variation evaluated separately. Augmentation independently varies response noise, baseline drift, timing, and channel masks within declared bounds. Force augmentation is paired with a compatible response transformation. Hyperparameter search must use the same budget across learned baselines, and both selected and rejected settings should be archived. This procedure limits selection bias; field generalization requires independent measured passages. Exact optimizer settings, window length, temporal-encoder specification, and graph-neighborhood limits remain necessary implementation details for independent reproduction.

4.3. Inference checks and implementation boundaries

At inference, the system validates passage metadata, loads the compatible calibration file, assembles the graph, and predicts one force per detected axle. The gross vehicle estimate is computed directly by summing the predicted axle forces, preserving the relationship between axle-level and vehicle-level errors. Predicted responses are reconstructed in dimensional units for comparison with the observed envelope. A large residual, poor timing confidence, extensive masking, or an unsupported axle arrangement produces a quality flag. Transferring a response residual threshold between bridges requires adjustment for channel sensitivity, filtering, and acquisition noise. An ill conditioned inverse operator can admit several axle allocations with similarly low residuals.

The implementation separates coordinate evaluation, graph construction, network execution, and response checking into individually traceable operations. Computational claims require a separate implementation benchmark with disclosed hardware, batch size, precision, and preprocessing scope. Sparse graph construction should be timed together with network execution because it can become significant for long passages. End-to-end timing includes coordinate evaluation and quality checks alongside accelerator execution. Deployment should also preserve the version of the calibration, preprocessing, graph rules, and learned weights as a single compatible configuration. An update to one component can change the distribution of node features received by the others. This dependency is particularly relevant after sensor replacement or a change in bridge support conditions, when a numerically valid influence file may nevertheless be incompatible with the calibration conditions represented during training. Acquisition latency depends on the response window required for axle separation, while computational latency begins once that window and its metadata are available.

5. Experimental Protocol and Data Analysis

5.1. Benchmark specification and matched comparisons

The evaluation compares prediction accuracy, component contributions, sensing robustness, and processing requirements under a common benchmark specification. The benchmark specifies 12000 vehicle passages, divided into 7200 training, 2400 validation, and 2400 test vehicles, with balanced groups containing two through six axles. Nominal axle forces span 40–160 kN, and five nominal speeds cover 10–30 m/s. A 30 m reference span carries six strain channels

sampled at 500 Hz. Load-cell-based bridge weighing has been evaluated through full-scale field tests and comparisons between estimated and reference vehicle weights [16]. Accordingly, calibration passages would form an independent dataset, with their own load measurements and timing records. The numerical exhibits present aggregate errors alongside distributional and condition specific diagnostics. Data-centric benchmarking compares neural architectures under controlled changes in sequence length, noise, frequency, and temporal delay [17]. Each comparator therefore receives the same response windows, trajectory inputs, masks, and partition definitions. Vehicle level blocking preserves dependence among the axles and windows of a passage. The primary metric averages axle absolute percentage errors within each vehicle before averaging across vehicles, while gross weight error evaluates the sum of the predicted axle forces. This distinction preserves the engineering importance of load allocation within a vehicle. All windows and axle labels belonging to one vehicle remain in the same partition, so the stated sample counts refer to independent passage records rather than window counts.

Figure 3(a) compares mean axle errors of 8.70%, 5.80%, 5.10%, 6.10%, and 3.20% for regularized inversion, temporal convolution, graph convolution, unencoded attention, and influence encoded attention, respectively. The 2.90 percentage point difference between the two attention configurations is examined alongside the feature ablations to assess the contribution of influence encoding. Figure 3(b) displays the corresponding error distributions, allowing the 5% reference threshold to be assessed against the distribution width and upper tail. A lower central value accompanied by a persistent tail would imply a need to identify specific failure conditions. Figure 3(c) relates load estimates to reference forces across 40–160 kN, with an identity line and a ±5% band. The widening absolute tolerance at higher forces makes relative agreement visible while retaining the dimensional magnitude of each error. Figure 3(d) presents error quantiles, including an 8.10% 95th percentile for the encoded model. The separation between its 3.20% mean and upper quantile directs attention toward uncommon passages that aggregate metrics may conceal. Relative to the unencoded attention error, the reduction is approximately 47.5%; the corresponding absolute difference is 2.90 percentage points.

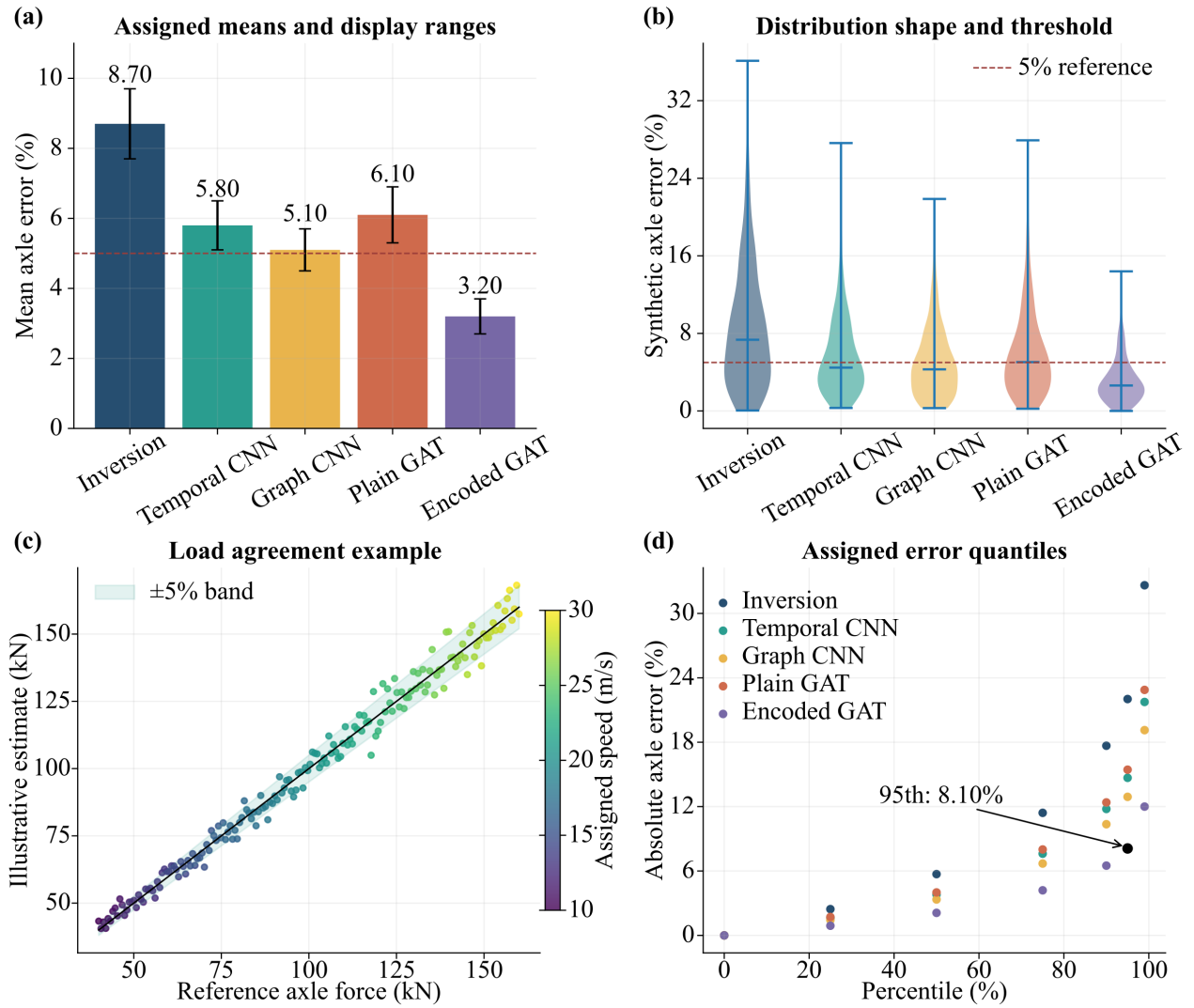


Figure 3. Baseline accuracy comparisons. (a) Mean axle error with declared display ranges. (b) Axle error violin distributions. (c) Load prediction scatter with identity and tolerance bands. (d) Absolute error quantile markers

Table 1 pairs these axle errors with gross weight errors of 4.90%, 3.20%, 2.80%, 3.40%, and 1.80%. The smaller gross errors are consistent with cancellation between opposing axle allocation errors; paired signed residuals would be required to establish the extent of that cancellation in measured data. Gross-weight accuracy therefore complements axle-level accuracy, while confirmation of cancellation requires paired signed errors for the individual axles within each vehicle.

Table 1. Baseline comparison of axle and gross weight errors

Method	Axle MAPE %	Gross MAPE %
Regularized inversion	8.7	4.9
Temporal convolution	5.8	3.2
Graph convolution	5.1	2.8
Graph attention	6.1	3.4
Influence encoded attention	3.2	1.8

Vehicle–bridge interaction analysis has examined how braking conditions, vehicle speed, and road roughness alter bridge dynamic responses [18]. Figure 4(a) compares mean errors of 2.80%, 2.90%, 3.10%, 3.40%, and 3.80% to speeds of 10, 15, 20, 25, and 30 m/s. The increase is modest at the lower levels and becomes more pronounced above 20 m/s, providing a defined pattern for investigating timing resolution and dynamic contamination. Figure 4(b) crosses the five speeds with five axle count classes, producing cell means from 2.20% to 4.40%. At each speed, the error increases by 1.20 percentage points between two-axle and six-axle vehicles. This pattern represents a hypothesized interaction between response overlap and the number of unknown loads. Coupled vehicle–bridge models exhibit sensitivity to vehicle speed, road roughness, damping, and suspension properties [19]. For the proposed experiment, identical signals would therefore be processed using reference trajectories, perturbed trajectories, and passage average speed. The speed trend would then be examined separately within each axle count to control for differences in vehicle class composition. Balanced speed means recover the 3.20% headline value, while the highest cell remains relevant for applications dominated by fast, multi-axle traffic. Each evaluation cell should retain its vehicle count, signed allocation errors, and vehicle-block uncertainty interval so that sparse conditions remain visible. The speed-stratified pattern establishes an association within the benchmark; attribution to timing error or dynamic contamination requires the controlled trajectory comparisons described here.

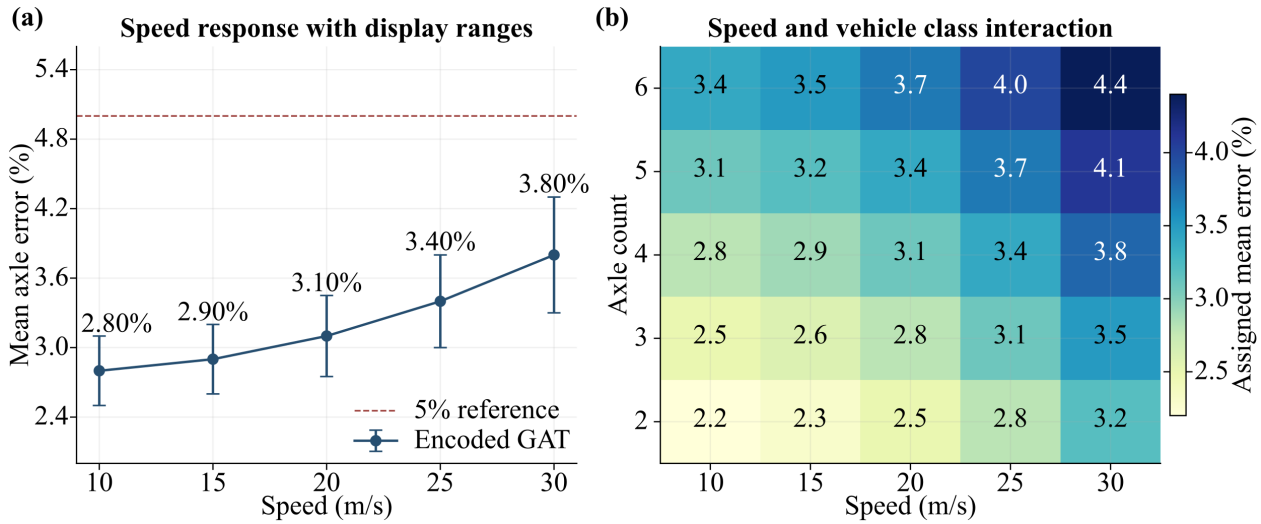


Figure 4. Operating condition sensitivity. (a) Speed dependent mean axle errors with ranges and a 5% reference line. (b) Mean error heatmap across five speeds and five axle count classes

5.2. Component attribution and robustness scenarios

Attention-based bridge monitoring has combined sensor, spatial, and channel representations within hierarchical networks for extreme-event detection [20]. This layered treatment motivates the following controlled comparison of model components. The variants preserve the temporal encoder and axle query head while removing ordinate features, slope features, response consistency, or moving coordinate edges. Figure 5(a) compares mean errors of 3.20%, 6.10%, 4.00%, 4.50%, and 4.80% to the full model and these four variants. Removing ordinates produces the largest increase, 2.90 percentage points, consistent with the intended role of position dependent sensitivity in distinguishing force magnitude from response amplitude. Removing slopes raises the error by 0.80 percentage points. That comparison

targets windows in which similar ordinates occur on different parts of an influence profile, where local gradient information could improve trajectory discrimination. Figure 5(b) shows validation learning trajectories ending at the same five validation values. The trajectories connect convergence behavior with terminal accuracy, showing that a smooth optimization path can terminate at a less accurate allocation rule. These one-component removals estimate effects conditional on the remaining architecture; interactions between slope information, moving edges, and consistency require joint ablations. Validation trajectories describe model selection and optimization behavior; final comparative accuracy must be assessed on the held-out test vehicles using the selected settings.

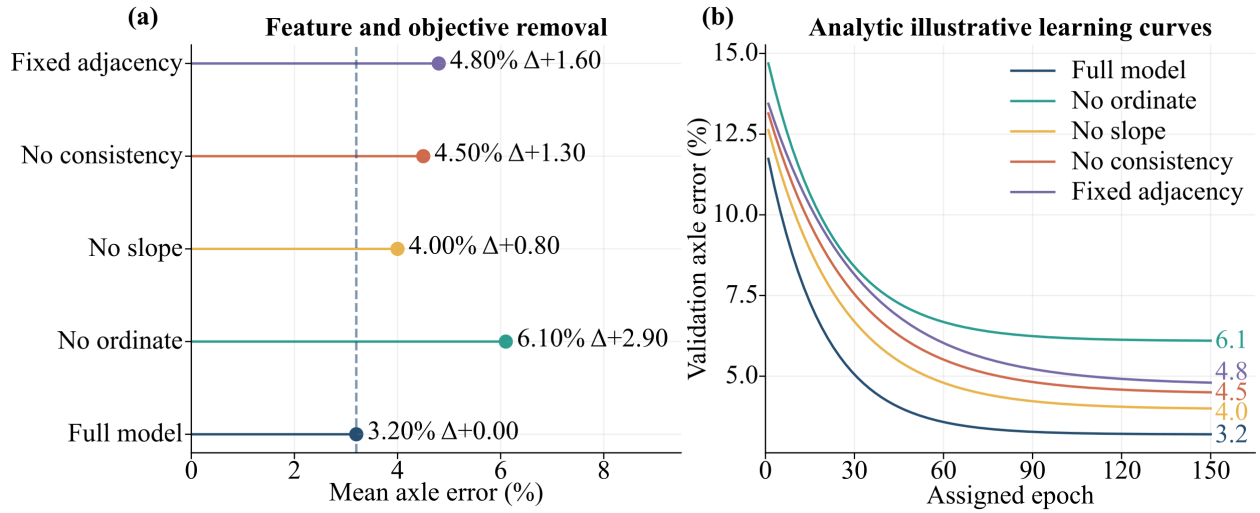


Figure 5. Component attribution. (a) Ablation errors in a lollipop chart. (b) Validation learning trajectories for five variants

Table 2 records the features removed from each variant so that architectural changes remain distinguishable from changes in input information. Reliable attribution requires repeated training under paired random seeds and an independently retained test partition, with parameter counts and stopping decisions archived for each variant.

Table 2. Component ablation comparison

Variant	Removed information	Axle MAPE %
Full model	None	3.2
No ordinate	Influence ordinate	6.1
No slope	Spatial derivative	4.0
No consistency	Response objective	4.5
Fixed adjacency	Moving edges	4.8

Bridge temperature prediction has combined physics-informed networks with parameter identification to represent thermal conditions and material interactions [21]. Here, influence calibration sensitivity is assessed through controlled perturbations of the response representation. Figure 6(a) relates axle errors of 3.20%, 3.40%, 3.80%, 4.50%, and 5.80% to normalized amplitude perturbations of 0%, 3%, 6%, 10%, and 15%. The final interval increases by 1.30 percentage points over a five-point perturbation increment, compared with an increase of 0.20 points over the initial three-point increment. This nonlinear trend directs the analysis toward a possible loss of tolerance as calibration mismatch increases. The 5% reference threshold is crossed between the last two sampled conditions; locating a practical boundary would require denser testing in that region. Figure 6(b) compares complete sensing, one missing channel, two missing channels, low noise, and high noise, with mean errors of 3.20%, 3.80%, 4.60%, 3.40%, and 5.10%. The box plots retain dispersion as well as the mean, since removal of an informative sensor could affect a subset of vehicle arrangements disproportionately. Comparisons of stochastic sensor-placement strategies examine how model uncertainty affects the information obtained from bridge monitoring layouts [22]. A random channel mask and loss of the most complementary measurement section should therefore be evaluated as separate scenarios. Their difference would reveal whether robustness depends primarily on redundancy or on preservation of distinct influence signatures. Calibration perturbation changes the encoded response operator, whereas observation noise alters the input signal; their error curves address different sources of degradation.

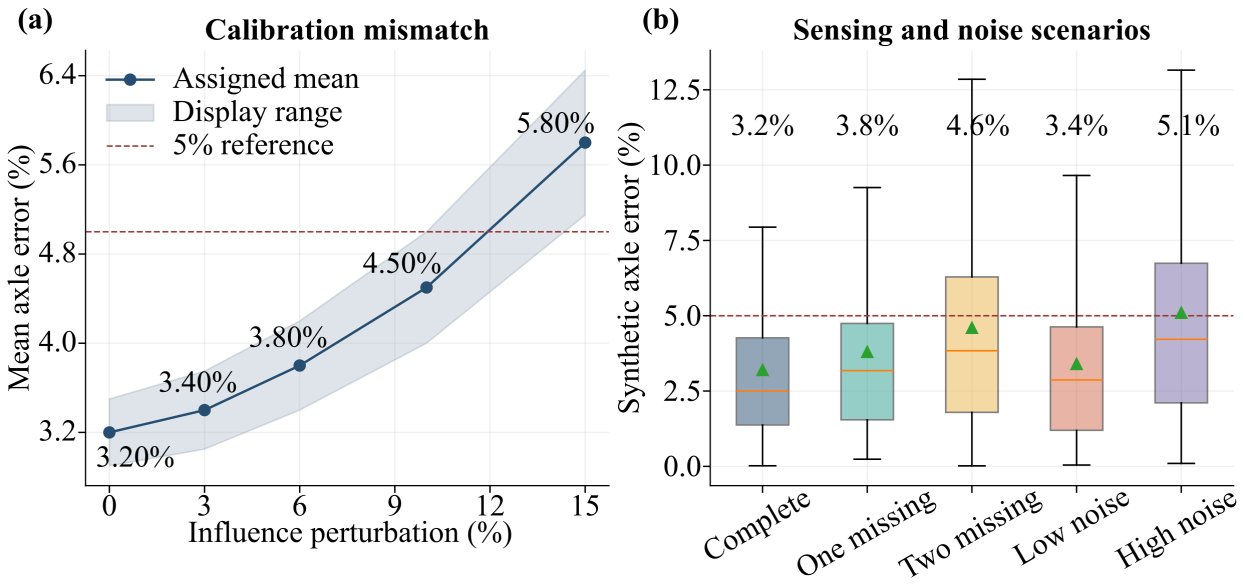


Figure 6. Robustness scenarios. (a) Influence perturbation sensitivity with reference threshold. (b) Error box plots for five sensing and noise conditions

The robustness scenarios involve two distinct error mechanisms. Acquisition noise changes the observations around a nominal response relation, whereas calibration mismatch changes the force-to-response relation itself. Railway vehicle–bridge interaction studies show that localized running-surface irregularities and train speed can substantially modify dynamic response [23]. The highway bridge reconstruction target similarly requires separate examination of dynamic residuals. A consistency penalty can suppress implausible predictions under accurate calibration, yet place excessive weight on a biased response operator after sensitivity drift. The ablation difference of 1.30 percentage points for consistency removal defines a benefit under the nominal condition; its persistence under increasing mismatch remains an experimental question. A factorial design would cross consistency weights with perturbation levels and record axle error, signed bias, and reconstruction residual for every passage. The same analysis should distinguish amplitude disturbance from coordinate shifts, because the latter can exchange information between adjacent axle queries. Filtering adds another identifiable tradeoff: attenuation of dynamic response can improve envelope reconstruction while weakening local features used for axle separation. Useful diagnostics would retain prefiltered response energy, filtered residual energy, timing confidence, and axle group spacing. Comparisons within matched speed and vehicle groups would then indicate whether deteriorating allocation follows reduced observation quality or a systematic change in the encoded sensitivity. Particular attention should be given to passages with a small reconstruction residual and a large axle error, as these expose ambiguity within the observation operator.

5.3. Residual structure and operational evaluation

Bland–Altman plots complement reliability indices by displaying systematic differences and agreement between measurements [24]. Figure 7(a) plots residuals against reference forces, with a zero line and force dependent ±5% bands. The band expands from ±2 kN at 40 kN to ±8 kN at 160 kN, making the distinction between relative tolerance and absolute force error explicit. Figure 7(b) displays differences against the mean of the reference and estimated loads, together with the mean bias and descriptive agreement limits. Clustering on one side of zero would indicate allocation bias, whereas increasing vertical spread would suggest load dependent dispersion. Figure 7(c) presents cumulative distributions for two- through six-axle groups, with mean errors of 2.70%, 2.90%, 3.20%, 3.40%, and 3.80%. These class diagnostics describe the subgroup assessment separately from the balanced aggregate comparison. Comparing their upper tails at the same error threshold tests whether the apparent advantage is concentrated among simpler vehicles. Figure 7(d) compares five sensor configurations through normalized axle error, retained channel count, and reconstruction residual. The error increases from 3.20% to 5.20% across the configuration sequence while the retained count declines from six to three. The intermediate five-channel configurations differ in error, emphasizing the role of sensor identity alongside channel count. A favorable reconstruction residual accompanied by degraded axle allocation would motivate examination of influence similarity within the retained channels. The relative tolerance bands express a selected error criterion, whereas the agreement limits summarize the distribution of signed differences; these visual references have different interpretations.

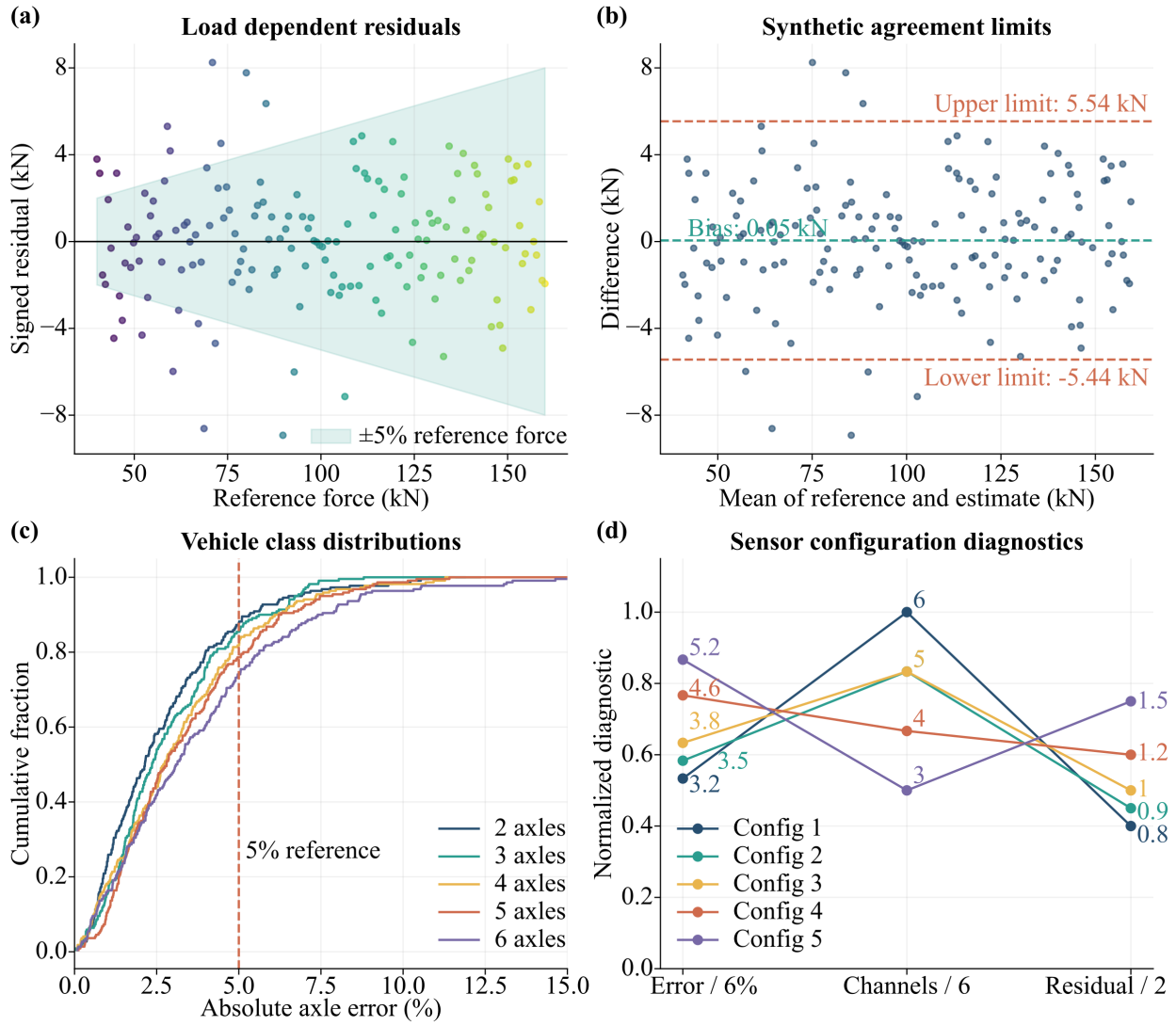


Figure 7. Residual and subgroup diagnostics. (a) Signed residual scatter. (b) Difference versus mean agreement plot. (c) Vehicle class empirical cumulative distributions. (d) Parallel coordinate sensor configuration diagnostics

Comparisons of cluster-robust, bootstrap, and jackknife inference show that uncertainty estimates can depend strongly on the inferential method [25]. A vehicle-level evaluation would resample vehicles and retain all their axle predictions, then compare paired method differences within each bootstrap replicate. Display ranges describe variation across the plotted conditions; population confidence intervals require vehicle-level paired outputs. Prediction uncertainty concerns a different target: the reliability of an interval for a particular axle. Conformal prediction research distinguishes marginal coverage from conditional guarantees and develops methods for coverage across specified subgroups [26]. For this framework, coverage should be assessed by axle count, speed, force level, and sensing configuration. A quality-based rejection policy would then report both the retained traffic fraction and error among retained passages. The rejected subset should retain its load distribution and acquisition flags, allowing failures involving sensor saturation, timing ambiguity, or weak influence sensitivity to be examined separately. Threshold selection belongs to an independent validation partition, with the final test evaluating the resulting fixed policy. This arrangement links operational acceptance to observable conditions. A model with a favorable retained-set mean but extensive rejection of heavy multi-axle vehicles would have a different monitoring value from one that maintains balanced coverage across the traffic population. Resampling should preserve complete vehicle records because errors from axles on the same passage share response conditions and are statistically dependent.

Studies of neural-network deployment evaluate inference performance across target hardware, execution settings, and software frameworks [27]. Figure 8 allocates 1.8 ms to signal preparation, 1.1 ms to coordinate evaluation, 2.4 ms to graph construction, 3.6 ms to inference, and 0.9 ms to checks. Their 9.8 ms total leaves 0.2 ms below the 10 ms budget. Inference accounts for approximately 36.7% of that total, while graph construction accounts for 24.5%; optimization confined to the network would therefore address only part of the processing chain. The stage breakdown identifies the operations that contribute most strongly to the total processing budget. Field bridge-weighing data combined with camera-based axle verification can improve reliability screening of vehicle-weight estimates [28]. The

corresponding study should pair static axle measurements with crossings at controlled speeds and lane positions, retaining reference uncertainty and environmental records. Processing time should be measured from availability of the required response window, while passage duration is reported separately. Long vehicles and slower crossings can increase acquisition latency even when computational latency remains stable. Sensor replacement should trigger repeat calibration checks and timing measurements, since changes in channel sensitivity can affect both the graph features and the frequency of quality flags. The benchmark record should retain preprocessing, calibration, graph, and weight versions for each timed configuration. The remaining margin is 2% of the displayed budget, so scheduling variability and hardware-specific overhead require separate timing measurements before operational use.

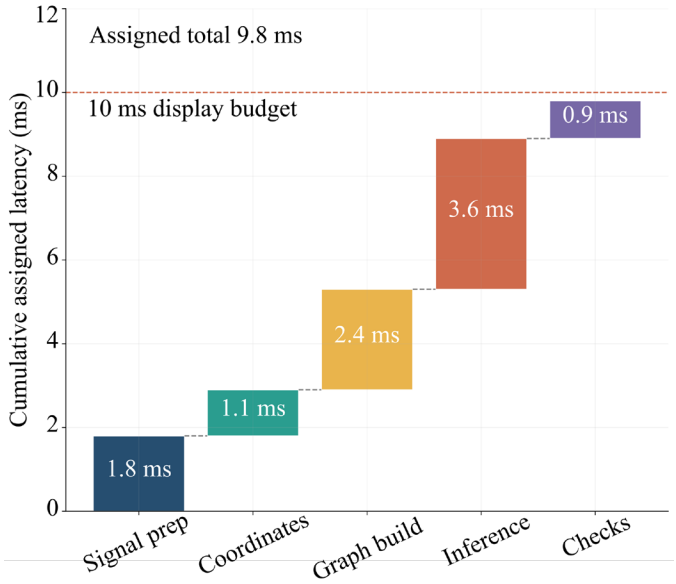


Figure 8. Processing latency waterfall for five stages with stage values, a 9.8 ms total, and a 10 ms display budget

6. Conclusion

This paper formulated a bridge weigh in motion axle identification framework that combines moving load influence encoding with graph attention. The observation representation links sensor windows to axle trajectories through calibrated ordinates, local slopes, and validity masks. The resulting graph organizes temporal continuity and complementary structural sensitivity, while axle query pooling produces positive force estimates for variable vehicle configurations. A soft response consistency objective provides an additional connection between predicted loads and the measured strain envelope. The manuscript also established a transparent evaluation design covering baseline comparisons, operating conditions, component removal, calibration mismatch, missing channels, residual structure, and processing requirements. The evaluation connects aggregate accuracy with axle allocation, calibration sensitivity, and sensing configuration.

The main limitations concern the dependence on bridge-specific calibration and the restricted representation of changing traffic and structural conditions. Establishing transferability requires assessment across distinct bridges, vehicle populations, and environmental states. The formulation depends on externally estimated axle timing and a compatible influence calibration, both of which may carry structured errors. Static response consistency only approximates vehicle bridge interaction and may bias force allocation when dynamic response or boundary changes are substantial. Attention weights describe model allocation of information, while causal interpretation requires controlled interventions. Quality flags require separate uncertainty calibration, and overlapping vehicles or unreliable axle detection extend beyond the present input assumptions. Transfer to another bridge requires compatible influence calibration and independent assessment of whether the learned correction remains valid under the new response distribution.

Future research should archive training configurations and extend evaluation to controlled moving load tests and independent field trials. Calibration uncertainty, variable speed, lane position, and sensor degradation should be perturbed separately and jointly. Repeated reference weighings would permit measurement uncertainty to be distinguished from identification error. Further work should examine uncertainty calibrated rejection, adaptive influence updates, and joint trajectory refinement while preserving information boundaries between calibration and testing. Operational assessment should document the complete acquisition and inference chain and define acceptance criteria for the intended bridge monitoring application.

7. Declarations

7.1 Author Contributions

Simona Pečnik contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision, project administration, and funding acquisition. Alenka Kobal contributes to validation, analysis, investigation, data collection, draft preparation. Klemen Majcen contributes to conceptualization, draft preparation, manuscript editing, visualization. All authors have read and agreed with the manuscript before its submission and publication.

7.2 Funding and Acknowledgments

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

7.3 Institutional Review Board Statement

Not applicable.

7.4 Informed Consent Statement

Not applicable.

7.5 Date Availability

Not applicable.

7.6 Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

7.7 Declaration of generative AI and AI-assisted technologies

During the preparation of this manuscript, the authors utilized Grammarly for linguistic polishing, logical organization and draft revision. The generative AI tool was not involved in research design, data collection, data analysis, or the equation of core conclusions. All outputs generated by AI were reviewed, revised and verified manually by the authors. The authors take full responsibility for all content of this manuscript.

8. References

[1] Rocheti, E. O., & Bacurau, R. M. (2024). Weigh-in-motion systems review: Methods for axle and gross vehicle weight estimation. IEEE Access, 12, 134822–134836. https://doi.org/10.1109/access.2024.3461653

[2] Irawan, J., Kim, C.-W., Cantero, D., & Yokoyama, T. (2025). Revised virtual axle method for damage detection using bridge weigh-in-motion systems. Journal of Civil Structural Health Monitoring, 15(8), 3731–3742. https://doi.org/10.1007/s13349-025-01013-w

[3] Ren, J., Li, M., Chen, Y., Zhang, Y., & Sun, Z. (2024). Semi-analytical model of vehicle-pavement-continuous beam bridge coupled system. Applied Mathematical Modelling, 131, 159–189. https://doi.org/10.1016/j.apm.2024.03.031

[4] Yan, W., Ren, H., Luo, X., & Li, S. (2025). Hybrid-data-driven bridge weigh-in-motion technology using a two-level sequential artificial neural network. Computer-Aided Civil and Infrastructure Engineering, 40(20), 2992–3012. https://doi.org/10.1111/mice.13415

[5] Li, C., Wu, C., & Zhang, C. (2025). Wavelet denoising based on comprehensive index optimization and improved L2 regularization for load identification. Journal of Vibration and Control, 31(7–8), 1093–1107. https://doi.org/10.1177/10775463241239790

[6] Hussain, S., Hussain, S. M., Xin, Y., & Wang, Z.-C. (2025). Vehicle load identification based on bridge response using deep convolutional neural network. Journal of Asian Architecture and Building Engineering, 24(3), 1235–1252. https://doi.org/10.1080/13467581.2024.2328634

[7] Ren, Z., & Zhou, Y. (2026). On cross-attention-based graph neural networks for fault diagnosis using multi-sensor measurement. Structural Health Monitoring, 25(4), 2716–2738. https://doi.org/10.1177/14759217251327760

[8] Michaloglou, A., Papadimitriou, I., Gialampoukidis, I., Vrochidis, S., & Kompatsiaris, I. (2026). Physics-informed neural networks in materials modeling and design: A review. Archives of Computational Methods in Engineering, 33(4), 5223–5260. https://doi.org/10.1007/s11831-025-10448-9

[9] Li, X., & Zhu, J. (2025). Identification of bridge influence line and multiple-vehicle loads based on physics-informed neural networks. Structural Health Monitoring, 24(2), 1167-1186. https://doi.org/10.1177/14759217241248570

[10] Hongqiu, L. I., Jiang, J., & Mohamed, M. S. (2026). Research on Dynamic Load Identification Based on the Newmark-β Method and Novel Iterative Regularization Strategy. Mechanics, 32(2), 107-115. https://doi.org/10.5755/j02.mech.42732

[11] Liangou, T., Zilakos, I., & Anyfantis, K. N. (2025). D-optimal sensor placement for load identification in wind turbine rotor blades. Journal of Intelligent Material Systems and Structures, 36(6), 411–427. https://doi.org/10.1177/1045389x251320802

[12] Dibiantara, D. P., Adha, A., Darmawan, M. S., Imjai, T., Russell, J., & Laory, I. (2025). Spatio–temporal enhanced anomaly detection in FRP bridge monitoring using MPCA, biGRU, and attention mechanisms. Journal of Civil Structural Health Monitoring, 15(6), 1837–1855. https://doi.org/10.1007/s13349-025-00913-1

[13] Masciotta, M. G., Barontini, A., Pellegrini, D., Brando, G., & Lourenço, P. B. (2024). Optimal sensor placement for bridge structural health monitoring: Integration of physics-based models with data-driven approaches. Procedia Structural Integrity, 62, 932–939. https://doi.org/10.1016/j.prostr.2024.09.125

[14] Mardani, K., Vretos, N., & Daras, P. (2024). Mahalanobis distance-based graph attention networks. IEEE Access, 12, 166923–166935. https://doi.org/10.1109/access.2024.3495531

[15] Shen, S., & Málaga-Chuquitaype, C. (2026). Learning rocking dynamics from sparse shake-table data with interpretable physics-informed neural networks. Earthquake Engineering & Structural Dynamics, 55(12), 3115–3136. https://doi.org/10.1002/eqe.70239

[16] Shi, S., Yarnold, M., Hurlebaus, S., & Mander, J. (2025). Evaluation of a load cell-based bridge weigh-in-motion system. Transportation Research Record: Journal of the Transportation Research Board, 2679(2), 535–549. https://doi.org/10.1177/03611981241263336

[17] Schlieper, P., Dombrowski, M., Nguyen, A., Zanca, D., & Eskofier, B. (2024). Data-centric benchmarking of neural network architectures for the univariate time series forecasting task. Forecasting, 6(3), 718–747. https://doi.org/10.3390/forecast6030037

[18] Wang, Y., Li, S., & Wei, N. (2024). Dynamic response impact of vehicle braking on simply supported beam bridges with corrugated steel webs based on vehicle-bridge coupled vibration analysis. Computer Modeling in Engineering & Sciences, 139(3), 3467–3493. https://doi.org/10.32604/cmes.2024.046454

[19] Mo, S., Chen, K., Huang, Z., & Zhang, W. (2024). Dynamic analysis and optimization of vehicle-bridge interaction system under road roughness and time variability of element interpolation function. Transportation Research Record: Journal of the Transportation Research Board, 2678(8), 294–309. https://doi.org/10.1177/03611981231215336

[20] Pan, Q., & Bao, Y. (2026). A hybrid attention hierarchical network-based extreme event detection method for structural health monitoring. Structural Health Monitoring, 25(6), 3882–3901. https://doi.org/10.1177/14759217251322943

[21] Wang, Y., Yang, D., Yuan, Y., Zhang, J., & Au, F. T. K. (2025). Hybrid physics-informed neural network with parametric identification for modeling bridge temperature distribution. Computer-Aided Civil and Infrastructure Engineering, 40(22), 3503–3524. https://doi.org/10.1111/mice.13436

[22] Barontini, A., Masciotta, M. G., Lourenço, P. B., & Brando, G. (2026). Compare stochastic optimal sensor placement strategies for bridge monitoring. Procedia Structural Integrity, 84, 352–359. https://doi.org/10.1016/j.prostr.2026.06.046

[23] Naseri, R., Mohammadzadeh, S., & Rizos, D. C. (2025). Rail surface spot irregularity effects in vehicle-track interaction simulations of train-track-bridge interaction. Journal of Vibration and Control, 31(5–6), 738–752. https://doi.org/10.1177/10775463241232024

[24] Barr, B. L., McIntosh, V. V. W., Britt, E. F., Jordan, J., & Carter, J. D. (2024). Visualizing agreement: Bland–Altman plots as a supplement to inter-rater reliability indices. Measurement: Interdisciplinary Research and Perspectives, 22(2), 175–187. https://doi.org/10.1080/15366367.2023.2246114

[25] Hansen, B. E. (2025). Standard errors for difference-in-difference regression. Journal of Applied Econometrics, 40(3), 291–309. https://doi.org/10.1002/jae.3110

[26] Gibbs, I., Cherian, J. J., & Candès, E. J. (2025). Conformal prediction with conditional guarantees. Journal of the Royal Statistical Society Series B: Statistical Methodology, 87(4), 1100–1126. https://doi.org/10.1093/jrsssb/qkaf008

[27] Alibekov, M. R., Berezina, N. E., Vasiliev, E. P., Vikhrev, I. B., Kamelina, Y. D., Kustikova, V. D., Maslova, Z. A., Mukhin, I. S., Sidorova, A. K., & Suchkov, V. N. (2024). Performance analysis methodology of deep neural networks inference on the example of an image classification problem. Numerical Methods and Programming (Vychislitel'nye Metody i Programmirovanie), 25(2), 127–141. https://doi.org/10.26089/nummet.v25r211

[28] Šoberl, D., Kalin, J., Anžlin, A., Kreslin, M., Čopič Pucihar, K., Kljun, M., Hekič, D., & Žnidarič, A. (2025). Enhanced precision in axle configuration inference for bridge weigh-in-motion systems using computer vision and deep learning. Computer-Aided

Civil and Infrastructure Engineering, 40(30), 6201–6216. https://doi.org/10.1111/mice.70144