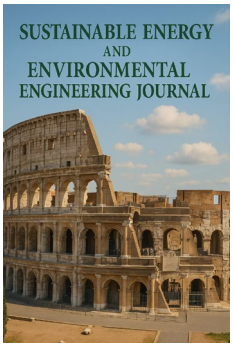
UNIVERSITY OF WEST ATTICA
eISSN:3136-8174

Sustainable Energy and Environmental Engineering Journal

Vol. 4, No. 1, January, 2026



DeepONet–FNO Estimation of a Road Dust Resuspension Intensity Index from Vehicle Trajectory Forcing and Particulate Sensor Data

Ladislav Drápal[1,*], Lubomír Hájek[1]

[1] *Department of Environmental Engineering, Faculty of Civil Engineering, Czech Technical University in Prague, Prague, 16627, Czechia*

Received 08 December 2025 Revised 14 March 2026; Accepted 20 March 2026; Published 23 March 2026

Abstract

Road dust resuspension estimation requires a distinction between vehicle disturbance, road surface state, and the transported particulate signal. This study evaluates a compact DeepONet–FNO configuration that combines learned branch coefficients on a fixed coordinate basis with a trainable spectral pathway to reconstruct a dimensionless spatial intensity index. Six input channels represent trajectory-derived forcing, particulate observations, and surface and transport descriptors. The evaluation comprises 720 training, 120 validation, and 180 independent test scenarios, each sampled at 32 spatial locations. The hybrid achieves a test root mean square error of 1.066 index points and a coefficient of determination of 0.9959, compared with 1.804 points and 0.9883 for the spectral comparator. These results describe the specified implementations and information representations. Increasing particulate noise from zero to 0.12 raises hybrid error from 0.989 to 1.552 points. Random removal and interpolation of half the particulate samples yields an error of 1.169 points while the other channels remain available. A joint moisture and transport shift reverses the primary ranking, with the spectral comparator outperforming the hybrid. Empirical residual bands calibrated on validation data attain 89.46 percent pooled pointwise coverage at a nominal level of 90 percent. The findings support further development of combined global and spatial representations under explicit observation assumptions. Physical calibration, independent field evaluation, and assessment of conditional reliability are required to establish their applicability to monitored road segments.

Keywords: Road Dust Resuspension, Vehicle Trajectory Fusion, Particulate Sensor Inversion, Deep Operator Networks, Fourier Neural Operators

1. Introduction

Road emission inventories describe traffic resuspension through relationships among surface loading, vehicle characteristics, and the material available for entrainment [1]. Near road observational studies distinguish locally generated particles from the atmospheric background entering a monitored corridor [2]. Investigations of deposited material indicate that the availability of loose particles changes with surface replenishment and removal, creating source variability beyond the instantaneous number of passing vehicles [3]. Research on non-exhaust source attribution consequently interprets a roadside concentration as a mixture shaped by several emission and transport processes [4]. Street scale transport studies describe the displacement and spreading that separate a source location from a concentration maximum [5]. These findings create a practical estimation problem: road maintenance acts on a particular segment, whereas a roadside sensor records particles transported from a wider area. Linking these scales requires an estimator that represents the location of vehicle forcing, the state of the road surface, and the transfer between source

* Corresponding author: ladislav.d@fsv.cvut.cz

https://doi.org/10.68406/seeej.2026.vol4iss1nz7:83-98

➢ 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.

and observation. Its output must also have a clearly defined physical interpretation, whether a measured mass flux, an emission factor, or a calibrated intensity index. The present work examines an intensity index within a controlled evaluation framework, allowing the information contributed by trajectories and particulate observations to be assessed under explicit source and transport assumptions. This formulation provides a basis for studying the spatial reconstruction problem before the additional uncertainty of field calibration is introduced.

Vehicle trajectories provide a more detailed description of road activity than a single count aggregated over a long interval. Trajectory reconstruction research has developed ways of recovering vehicle motion from incomplete position records, making local activity profiles available for environmental inference [6]. Studies of tire–road particle formation have examined how driving conditions and the interaction of tires with road material affect observed particle characteristics [7]. Sensor calibration investigations document the sensitivity of low-cost particulate observations to instrument response and environmental conditions [8]. Research on data assimilation has addressed the complementary problem of reconciling imperfect observations with a dynamical description of an environmental field [9]. Inverse source studies show why source recovery becomes unstable when transport removes spatial detail or observations constrain only part of the contributing processes [10]. Combining these approaches requires an explicit account of their different physical meanings. A trajectory sequence describes forcing activity, while concentration records a transported particle mixture. Spatial interpolation produces a complete sampling grid whose information content remains determined by the original measurements. A model can appear accurate if adjacent samples from the same event are divided between training and evaluation. The estimation problem therefore requires an explicit observation model, scenario level separation, and tests that distinguish reduced interpolation error from recovery of physically meaningful source structure. These requirements are particularly relevant when engineering studies introduce flexible neural representations with more capacity than the available independent environmental episodes can support.

Operator learning offers a way to describe the required mapping as a transformation between spatially distributed inputs and an output field. In the proposed architecture, a compact branch pathway encodes scenario context and combines it with coordinate basis functions, while a spectral pathway preserves spatially varying information through local channel transformations and Fourier mixing. Their combination is intended to address two different requirements: conditioning on the overall traffic and road state, and retaining the locations of local forcing and measured particulate variation. Their complementarity depends on the operating conditions and the information retained by each representation. A global summary can discard event location, and a spectral transformation can introduce periodic boundary assumptions requiring modification for an open road. The study consequently evaluates a compact implementation in a fully specified evaluation framework, using a pointwise linear comparator and separately trained pathway comparators. Its contribution is an evaluation of a compact fusion configuration using shared data, saved weights, residual diagnostics, and stress tests. The comparison assesses the documented input representations and capacities; it motivates complementarity without isolating an architectural mechanism. The experimental output is a dimensionless index expressed on a scale of one hundred points per reference intensity unit. Its interpretation is restricted to the reference intensity scale; conversion to a physical emission factor requires external calibration. This restricted scope permits the central architectural question to be examined while leaving field identifiability as a separate empirical requirement.

Establishing a credible connection between traffic activity and inferred road dust intensity requires a clear progression from the observation assumptions to the model formulation and the evidence used to assess reconstruction quality. Section 2 reviews road dust source estimation and the operator representations relevant to multimodal inverse problems. Section 3 defines the intensity target, the six input channels, and the observation mechanism, including the assumptions that make the target observable in the benchmark. Section 4 specifies the hybrid architecture, the nonnegative output transformation, the optimization objective, and the evaluation protocol. Section 5 presents the evaluation dataset and analyses of reconstruction accuracy, residual structure, missing observations, noise, input sensitivity, calibration, and distribution shift. It reports failures alongside improvements so that the evidence can be interpreted at the level supported by the implemented comparators. Section 6 identifies the methodological outcome, the remaining limitations, and the field and computational investigations required for further development. All experimental claims refer to the reproducible evaluation framework described in the paper. This organization keeps the observation assumptions visible before model evaluation and places the most detailed numerical analysis after the architecture has been specified and its scope established. The resulting sequence connects implementation choices with the conditions needed for interpreting the reported estimates.

2. Related Work

2.1. Road dust sources and particulate observation

Engineering approaches to road dust estimation often begin with relationships between road surface properties, vehicle activity, and the amount of material available for entrainment. Integrated traffic-emission estimation and source attribution require consistent treatment of averaging scales and the transfer between emission and observation [11]. Such

approaches are useful comparators when their intended averaging period and source definition are preserved. Interpreting a segment scale estimate alongside a short concentration excursion requires an explicit representation of the transfer between emission and observation. Mechanistic interpretations introduce additional state variables, including surface wetness, replenishment, and the fraction of material susceptible to disturbance. Mechanically induced resuspension studies emphasize the availability of deposited material and motivate explicit treatment of the surface state [12]. A supervised model may approximate this state implicitly, while unique source attribution requires additional information about the history of deposited material. The present benchmark therefore exposes loading and moisture as observable inputs to define the state information available to the network.

Observation driven methods attempt to exploit the temporal association between traffic and measured particles. Research combining roadside measurements with traffic information has investigated how transport and source composition alter that association [13]. The value of trajectory information lies in its representation of when and where forcing occurs; separating resuspended material from other particulate contributions remains an additional requirement. Instrument studies also show that particle measurements require attention to calibration and environmental response, especially when data are collected by compact sensors [14]. An observation pipeline should retain missingness and calibration metadata even if a learning model consumes a regular grid. In this study, the numerical concentration is a blurred, noisy transformation of a known source index. Representing exhaust, brake wear, humidity induced optical bias, and independently varying background concentrations would require additional source and observation components. This simplified observation model permits controlled attribution of reconstruction errors to the specified architecture and perturbations; field evaluation requires a broader representation of particle sources and sensor responses. Evaluation of source separation would require additional measurements and an experimental design that explicitly represents competing sources.

2.2. Operator representations and multimodal inverse estimation

Deep operator representations separate an encoding of an input function from an encoding of the coordinates at which an output is queried. Work on branch and trunk constructions establishes a general framework for learning mappings between functions through separate representations of input functions and output coordinates [15]. The information retained by a finite encoder determines the inverse tasks it can represent. A branch formed from spatial averages can represent scenario context efficiently while losing the location of a narrow emission event. Spectral operator architectures address another aspect of the problem by learning transformations of spatial frequency coefficients and combining them with pointwise channel mixing [16]. Their global receptive field is attractive for transport affected observations, although mode truncation and the treatment of boundaries remain consequential implementation decisions. The compact model used here retains a limited set of modes and uses a fixed coordinate basis. Its scope is a constrained instance of operator learning with a specified discretization and sensor layout.

Multimodal inverse learning adds the challenge of combining inputs that differ in physical interpretation and information content. Research on learned inverse operators emphasizes the role of priors and observation operators in determining which parts of an inferred field are data supported [17]. A trajectory channel can act as a strong prior on source location, while concentration provides a second constraint that reflects transport. If the same deterministic rule generates both the labels and an informative input, high reconstruction accuracy can reflect access to that rule under the assumptions of the source formulation. Uncertainty research in environmental machine learning has consequently distinguished interval calibration under a sampled distribution from reliability under a changed distribution [18]. This distinction guides the numerical protocol: an ordinary held out set assesses interpolation among evaluation scenarios, and a separate wetness and transport shift tests one extrapolative condition. Residual intervals are calibrated only on validation scenarios and reported as marginal pointwise intervals. Their interpretation is marginal and pointwise, with simultaneous spatial coverage and transfer calibration requiring separate evaluation. These definitions keep the architectural comparison aligned with the evidence available from the evaluation.

3. Problem Formulation and Observation Model

3.1. Intensity target and trajectory representation

The computational domain contains 32 equally spaced spatial locations from zero through 31/32 in normalized road length. The unit endpoint is excluded to avoid duplicating the periodic boundary. Each scenario represents an independent traffic and road state snapshot defined by the evaluation framework. Between three and seven localized vehicle forcing events are sampled for each scenario. Each event contributes a Gaussian spatial footprint with a random center, width, and amplitude. Their spatial superposition defines the trajectory forcing channel through the following equation:

$$q(x) = \frac{1}{3}\sum_{j=1}^{J} a_j exp\left[-\frac{(x - x_j)^2}{2w_j^2}\right] \quad (1)$$

Here, J is the number of sampled events, while a_j, x_j, and w_j denote the amplitude, center, and width of each event footprint. This channel approximates the spatial distribution of vehicle disturbance activity. The forcing representation aggregates vehicle events, with explicit treatment of lane changes, vehicle wakes, and tire contact mechanics requiring a more detailed traffic model. The evaluation therefore concerns information fusion after trajectory aggregation, at the spatial operator stage of the sensing pipeline.

Footprint widths are drawn between 0.035 and 0.12 in normalized road length, and amplitudes between 0.3 and 1.3 before the sum of footprints is divided by three. The sampled events represent disturbance footprints rather than verified counts of individual vehicles. Field deployment would require a mapping from measured vehicle positions, speeds, classes, and timestamps to an equivalent forcing representation. This mapping must preserve physical units and represent the spatial and temporal support of each observation. The additional speed channel supplies a scenario multiplier independently of the sampled event positions. The numerical results consequently characterize estimation from aggregated forcing; assessment of trajectory extraction requires a separate evaluation against measured vehicle motion. Sensitivity to speed measures the response imposed by the specified speed exponent.

3.2. Surface state and particulate transfer

The source formulation combines local forcing with scenario level loading, speed, and moisture factors. A fixed spatial modulation introduces a mild coordinate dependence that requires positional information alongside local channel values. Loading is sampled between 0.4 and 1.6, speed between 0.65 and 1.35, and moisture between 0.05 and 0.55. All three are dimensionless variables in the source formulation. Their exponents and modulation coefficients prescribe the response represented by the source equation; physical calibration of this response requires an independent source reference. The source-generation equation expresses these dependencies as

$$s(x) = [0.10 + q(x)^{1.25}]LV^{1.4}exp(-2M)[1 + 0.18sin(2\pi x)] \quad (2)$$

The variables L, V, and M represent loading, speed, and moisture, respectively. The output displayed in the experiments is the reference source multiplied by one hundred, with the resulting unit termed an index point. The source remains positive where trajectory forcing is small because the source formulation includes a background disturbance term. This positive offset belongs to the source index definition. A separately varying atmospheric background would enter through an additional observation component.

The source is transformed into a particulate observation by a seven-tap kernel with one-directional spatial ordering. Causality here refers to that ordering rather than to elapsed-time transport. Its integrated gain varies with the transport descriptor instead of being normalized to unity. A transport variable between 0.6 and 1.5 controls the decay length. Noise is added after transfer using independent Gaussian draws with a standard deviation of 0.04 for the primary experiment. The transfer is evaluated with circular shifts, so material leaving the right boundary reappears at the left. Application to a finite open road requires boundary conditions representing particle inflow and outflow at the monitored segment limits. These operations give the observation equation

$$c_i = \sum_{\ell=0}^{6} \frac{exp[-\ell/(1.5 + U)]}{3} s_{(i-\ell) mod 32} + \varepsilon_i \quad (3)$$

Here, c_i is the particulate proxy at grid location i, U is the transport descriptor, and ε_i is the additive observation noise. The kernel has a transport dependent integrated gain, allowing the concentration channel to carry both spreading and amplitude information. Its values are normalized sensor proxies on the reference scale. Figure 1 presents the complete generation and observation workflow, including event aggregation, surface modulation, transfer, noise, and scenario separation. The diagram distinguishes the latent target from the observed channels and identifies the target information available during training.

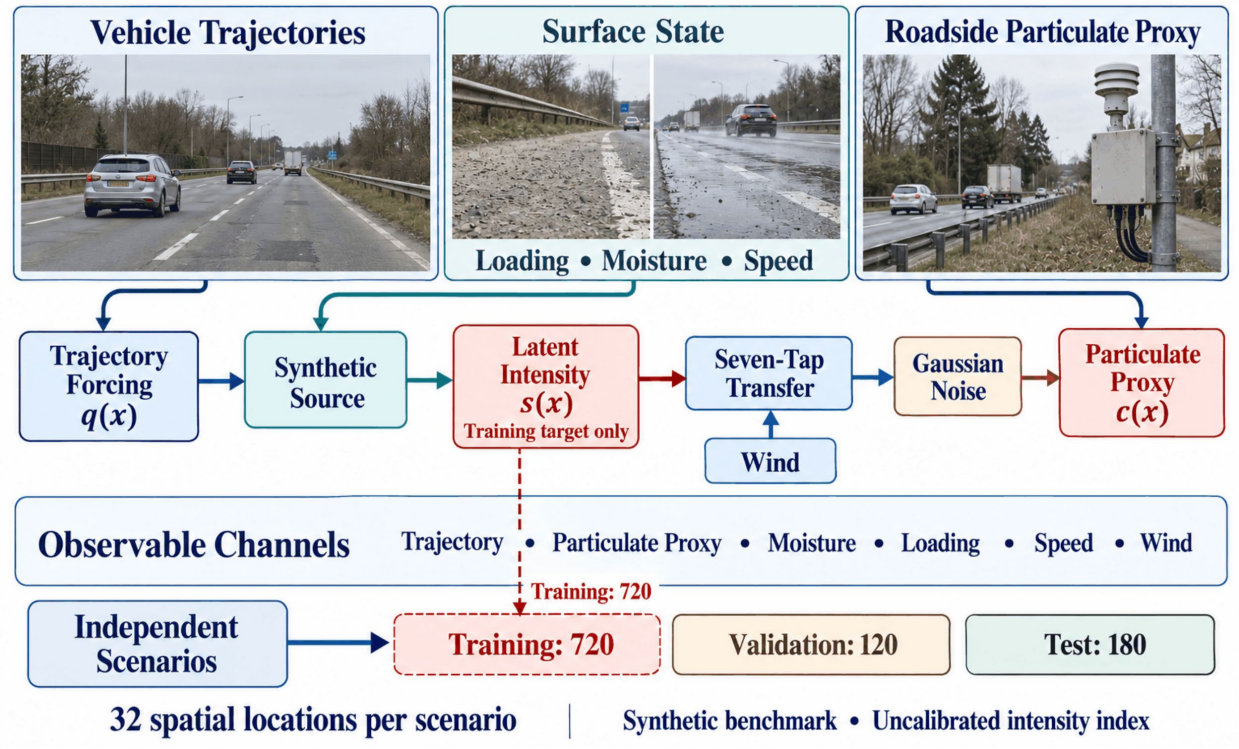


Figure 1. Workflow linking trajectory forcing, surface state, latent intensity, and noisy particulate observations before independent scenario splitting

3.3. Observable channels and identifiability

The input has six channels at every grid location: trajectory forcing, particulate proxy, moisture, loading, speed, and transport state. The last four are broadcast across the grid because they are scenario descriptors. This representation supplies the network with more road state information than many monitoring systems can measure routinely. It establishes a controlled reference problem in which the information available for estimation is explicitly specified. If loading is unknown, or moisture is inferred from a remote weather measurement, the corresponding uncertainty must be propagated through the input representation and the inferred intensity. An independently varying concentration background would require an additional input or latent state to distinguish transported background particles from locally generated material. Deployment requires synchronized trajectory aggregation, surface-state measurements or calibrated proxies, and an independent source reference for mapping the index to physical emissions. Uncertainty in those inputs must be propagated into the segment estimate.

Identifiability in this experiment is conditional on the evaluation framework and the supplied channels. The trajectory and road state channels constrain the source strongly, while concentration provides a noisy, spatially mixed observation of the same quantity. This formulation defines a fused supervised inverse surrogate conditional on the supplied forcing and surface information. Evaluation under a field observation regime could replace one or more directly supplied surface descriptors with measured proxies and propagate their uncertainty into the inferred intensity. The present study assesses fitted input dependence through channel permutation after training and evaluates observation sensitivity through separate perturbations of particulate noise and availability. These interventions reveal whether the fitted model uses the intended inputs, with interpretation restricted to fitted dependence. Independent scenario generation and a separate shifted test set distinguish interpolation within the specified domain from extrapolation beyond its moisture and transport ranges.

4. Hybrid Operator Architecture and Evaluation Protocol

4.1. Global context and spectral pathways

The branch pathway begins with the spatial mean and standard deviation of each of the six channels. These twelve descriptors pass through a trainable dense layer of width twelve with a hyperbolic tangent activation, followed by a linear transformation producing nine coefficients. The trunk consists of a constant and sine and cosine functions at four spatial frequencies. The coordinate functions remain fixed during optimization, while the branch learns their scenario-dependent coefficients. Combining the branch coefficients with the trunk basis yields a separable operator equation:

$$d(x) = \sum_{k=1}^{9} b_k(\boldsymbol{g}) t_k(x) \tag{4}$$

The descriptor vector **g** contains the twelve global input statistics, $b_k(\mathbf{g})$ denotes a learned branch coefficient, and $t_k(x)$ denotes a coordinate basis function. The resulting pathway is a compact branch and trunk operator of the DeepONet family. Its restricted encoder compresses event locations into global statistics, a property relevant to interpreting its standalone comparator. The DeepONet designation therefore identifies a fixed-basis branch–trunk variant in this implementation.

The spectral pathway lifts the six spatial channels to eight latent channels using a trainable affine transformation and a hyperbolic tangent. It then combines a pointwise channel transformation with a learned Fourier transformation retaining the first six nonnegative frequency bins. A second hyperbolic tangent precedes a scalar projection. The retained bins have indices zero through five, including the constant mode; the Nyquist bin at index sixteen is excluded. Discarded coefficients are set to zero. The reported comparison uses the fixed 32-point grid, and transfer across discretizations requires separate assessment. Complex spectral weights are represented by separate real and imaginary parameter arrays. The inverse transform returns a real spatial field, with orthonormal transform normalization applied consistently in the forward computation and its gradients.

$$\boldsymbol{z} = tanh\{\boldsymbol{W}\boldsymbol{h} + \mathcal{F}^{-1}[\boldsymbol{R}\mathcal{F}(\boldsymbol{h})] + \boldsymbol{a}\} \tag{5}$$

In this spectral update equation, **h** is the lifted latent field, **W** performs local channel mixing, **R** contains the learned transformations of the retained Fourier coefficients, and **a** is the latent bias. The operators $\mathcal{F}$ and $\mathcal{F}^{-1}$ denote the forward and inverse Fourier transforms. A learned scalar projection of **z** produces the spatial pathway output $f(x)$.

The implementation uses one spectral mixing block. This compact configuration supports reproduction of the model evaluation using NumPy and explicitly implemented gradients. Figure 2 shows the two pathways, their input representations, additive fusion, and positive output transformation. It also identifies the training connection from the target to the loss. The global pathway introduces coordinate dependent modulation, while the spectral pathway retains spatial forcing information. Additive fusion combines these contributions through a direct sum of the pathway outputs.

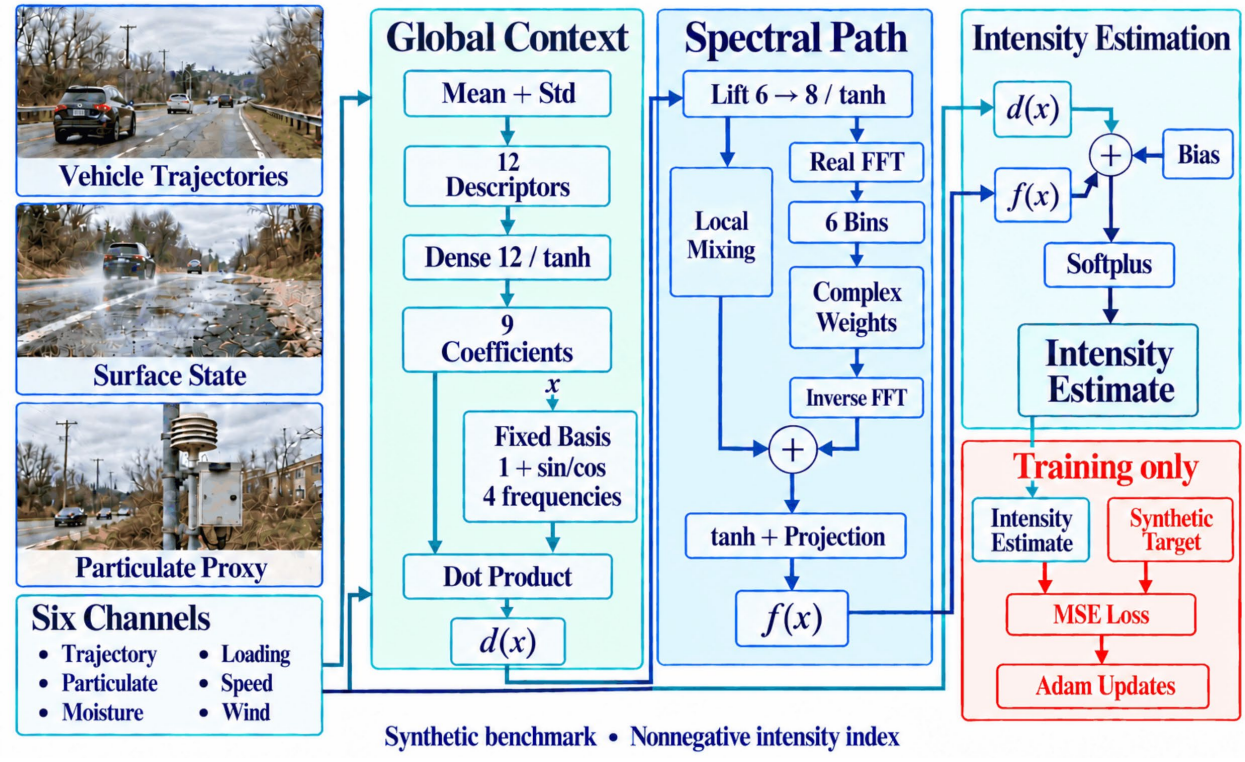


Figure 2. Compact DeepONet FNO architecture with a twelve-descriptor branch, nine coordinate basis functions, an eight-channel spectral block, six retained modes, and positive fused output

4.2. Optimization and comparator definitions

The two pathway outputs are added with a scalar bias and passed through a softplus transformation. Training minimizes mean squared error on the unscaled reference intensity. Multiplication by one hundred converts the reported errors to index points while preserving the optimization and model ordering. The output parameterization enforces

positivity, and the objective measures supervised reconstruction error against the reference source labels. The softplus equation defines the final estimate as

$$\hat{s}(x) = log\{1 + exp[d(x) + f(x) + b]\} \quad (6)$$

The scalar b is the fusion bias. The transformation produces a positive estimate while retaining a smooth dependence on both pathway outputs, allowing their parameters to be optimized jointly through the reconstruction objective.

Optimization uses Adam updates with a learning rate of 0.006, first and second moment coefficients of 0.9 and 0.999, and a numerical stabilizer of 10^{-8}. Each gradient component is clipped to the interval from minus one to one. Training lasts 180 epochs with shuffled minibatches of 96 scenarios, using the same deterministic initialization seed and shuffling seed for the neural comparators. The final epoch is retained; validation is reserved for error assessment and residual interval calibration. A training loss trace is recorded every ten epochs. The explicit gradient implementation supports execution in NumPy, with validation against a second implementation remaining a useful computational check. Each neural score describes one fitted realization; repeated initializations and scenario samples are needed to quantify variation in the ranking.

The comparators are a pointwise ridge regressor, the branch and trunk pathway alone, and the spectral pathway alone. Ridge receives the same six channels at each location, includes an intercept, and uses a regularization coefficient of 0.01. Its outputs are clipped at zero. The standalone neural pathways are independently initialized and optimized under the same training protocol. They share the optimization budget, while parameter counts and representational capacities vary with their inputs. The branch comparator uses global descriptors, whereas the spectral comparator receives the full spatial input. Their comparison evaluates the documented information pathways at the specified capacities. Equal training budgets leave capacity and spatial-information access as separate sources of variation. The branch-only score applies to the restricted global-statistics encoder used here.

4.3. Data separation and stress tests

The training, validation, and primary test sets contain 720, 120, and 180 independent evaluation scenarios. Their random seeds are 20261020, 20261021, and 20261022, respectively. All locations within a scenario remain in the same partition. The primary test contains 5760 point predictions, with the 180 scenarios constituting the independent experimental units. Root mean square error, mean absolute error, coefficient of determination, and signed mean error are computed over all point predictions. These pooled metrics weight each scenario equally because every scenario contains the same number of locations. They summarize the evaluation sample; scenario-level resampling is needed to quantify between-scenario sampling uncertainty. Sampling uncertainty across cities and measurement campaigns requires independently collected field episodes.

Noise sensitivity uses six standard deviations between zero and 0.12 while holding the test source realizations fixed. Missing observation tests replace independently removed particulate samples by one dimensional linear interpolation within each scenario. Interpolation extends the nearest available endpoint value at the domain edges, while the observation operator applies periodic transfer; the resulting error therefore reflects both particulate deletion and this preprocessing choice. Trajectory and surface channels remain complete, and structured or simultaneous channel outages require separate tests. Six missing fractions from zero to one half are examined using the fixed trained weights. Channel sensitivity is measured by permuting an entire channel between test scenarios, preserving its spatial structure within each scenario.

A separate test draws moisture between 0.55 and 0.80 and transport between 1.50 and 2.00. These ranges extend beyond the corresponding training conditions and provide a defined assessment of combined environmental shift. Residual intervals use absolute validation residual quantiles at six nominal levels. Their empirical test coverage is calculated pointwise, and their widths are reported in index points. The radii use pooled absolute validation residuals and linearly interpolated empirical quantiles. These are empirical pointwise residual bands; within-scenario dependence and the unadjusted quantile convention limit finite-sample coverage claims for entire spatial scenarios.

5. Model Evaluation and Analysis

5.1. Dataset characteristics and reconstruction accuracy

The experimental population comprises independent traffic and road state scenarios with a shared spatial observation mechanism. Environmental evaluation methodology treats the episode as the independent sampling unit when several correlated measurements describe the same event [19]. Accordingly, all 32 locations belonging to a scenario remain within one partition, producing 720 training, 120 validation, and 180 test scenarios. Table 1 specifies the event count, surface ranges, transport range, and observation noise used in these partitions. The primary test contains 5760 target values spanning 1.12 to 169.82 index points, with a median of 11.55 and a ninetieth percentile of 35.38. Overlapping forcing footprints and the multiplicative surface factors generate the upper tail. This population creates two distinct

reconstruction demands: resolving the numerous low intensity locations and retaining the less frequent peaks that contribute strongly to squared error. The fixed spatial modulation also makes the source response depend on position, providing a defined role for coordinate information. Because the surface descriptors are observed directly, the experiment evaluates approximation under known surface state. Field calibration would additionally require measured loading, moisture, and an independent reference for source intensity. The evaluation settings therefore define the interpretation of every reported error and the environmental conditions represented by the evaluation sample.

Table 1. Dataset settings and observation ranges

Quantity	Setting	Interpretation
Training scenarios	720	Seed 20261020
Validation scenarios	120	Seed 20261021
Test scenarios	180	Seed 20261022
Locations per scenario	32	Normalized periodic domain
Forcing events	3 to 7	Gaussian spatial footprints
Moisture M	0.05 to 0.55	Shift: 0.55 to 0.80
Loading L	0.40 to 1.60	Dimensionless
Speed V	0.65 to 1.35	Dimensionless
Transport U	0.60 to 1.50	Shift: 1.50 to 2.00
Particulate noise SD	0.04	Stress range 0.00 to 0.12
Intensity scale	100 times s	Dimensionless index points

Measurement agreement analysis examines the response range and the position of systematic deviations alongside pooled correlation [20]. Figure 3(a) compares the hybrid estimate with all test targets using equal axis scales, a parity reference, and a diagnostic corridor of five index points. The hybrid places 99.76 percent of predictions within this corridor and 92.74 percent within two points, indicating that most locations have small absolute deviations under the sampled observation regime. The sparse upper end of the scatter remains relevant because a small number of peak errors can affect the pooled RMSE disproportionately. Figure 3(b) examines spatial reconstruction in the first test scenario, whose target peak is 26.01 points. The hybrid, spectral, and branch comparators produce peaks of 27.24, 24.08, and 13.74 points, respectively. The hybrid preserves the prominent local rise while slightly exceeding its amplitude; the spectral estimate retains the peak location with a lower maximum. The branch comparator spreads the response over a broader spatial profile, consistent with the compression introduced by global input summaries. The first scenario is selected by its fixed index as an illustration rather than as evidence of representative peak behavior. The five-point corridor is a descriptive reference on the index scale, with operational acceptance thresholds requiring independent validation. Its peak behavior illustrates how the models distribute error spatially within one observation realization.

Figure 3(c) displays the intensity histogram, showing the concentration of test values below the upper tail and the separation between the median of 11.55 and the ninetieth percentile of 35.38 points. The resulting imbalance gives low intensity locations greater influence on metrics that average absolute deviations across points. Clustered prediction assessment distinguishes the number of residuals from the number of independent episodes contributing to those residuals [21]. The interpretation here therefore retains the 180 scenario sampling units when describing the 5760 plotted points. Figure 3(d) maps signed hybrid errors over the first twelve scenarios, exposing whether deviations persist across neighboring locations. Positive residuals correspond to overestimated intensity and negative residuals to underestimated intensity; contiguous regions of the same sign indicate a local displacement or amplitude error that a single pooled score can conceal. Spatially organized deviations are plausible because the observation kernel mixes neighboring source values and the coordinate pathway introduces a smooth position dependent contribution. A numerical check across four intensity bands gives hybrid RMSE values of 0.481, 0.965, 1.406, and 2.250 points for targets below 10, from 10 to 20, from 20 to 40, and at least 40 points, respectively. The increasing conditional error shows why the densely populated low range and the sparse upper range require separate interpretation.

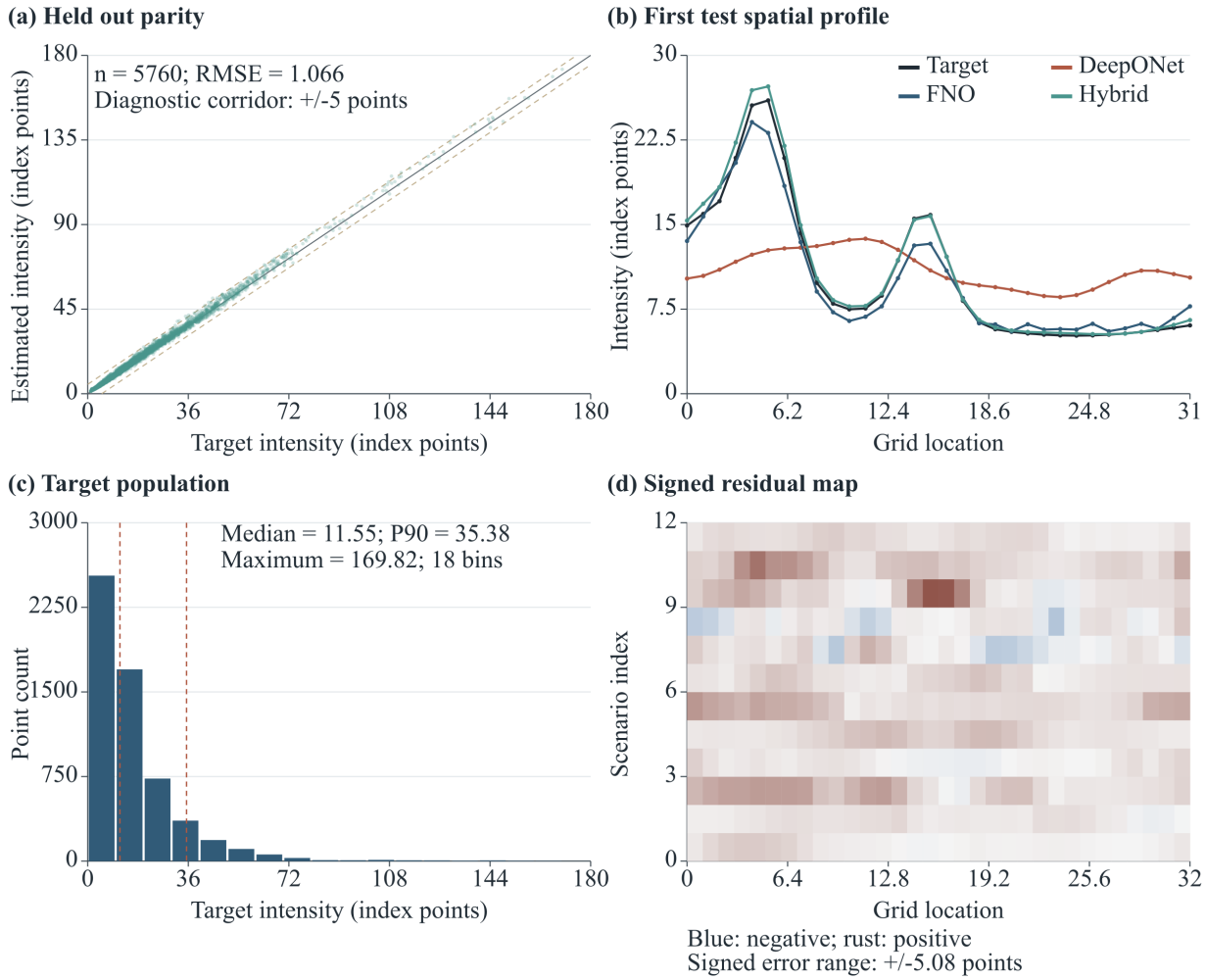


Figure 3. Held out reconstruction and target composition. (a) Hybrid parity scatter with a five-point diagnostic corridor. (b) First test scenario spatial profile. (c) Target intensity histogram with quantile markers. (d) Signed residual heatmap for twelve test scenarios

Table 2 provides a direct comparison of the primary test and validation errors. The hybrid attains an RMSE of 1.066 points, an MAE of 0.706, and an R squared value of 0.9959; the spectral comparator gives 1.804, 1.264, and 0.9883. Relative RMSE and MAE reductions are 40.92 and 44.15 percent within this benchmark, calculated as the error reduction divided by the spectral comparator error. These descriptive reductions characterize the fitted implementations rather than statistical significance or operational benefit. Ridge gives 5.301 and 3.604 points, while the branch comparator gives 11.391 and 7.493. The ridge result shows that a local linear mapping already explains a substantial fraction of variation when forcing and surface descriptors are available. The hybrid validation RMSE of 1.043 closely matches its test error, and the spectral validation error of 1.699 remains near its test value of 1.804. This agreement indicates similar approximation difficulty across independent samples from the evaluation dataset. Signed error introduces a separate distinction: the hybrid bias is positive at 0.509 points, compared with minus 0.357 for the spectral comparator. An integrated source index would accumulate this positive offset across road locations, even where individual errors are modest. The numerical advantage in reconstruction accuracy therefore accompanies a directional residual component that merits examination before an operational ranking or maintenance threshold is defined.

Table 2. Primary held out performance in index points except R squared

Model	RMSE	MAE	R squared	Bias	Validation RMSE
Ridge	5.301	3.604	0.8990	0.387	N/A
DeepONet	11.391	7.493	0.5335	-0.258	11.253
FNO	1.804	1.264	0.9883	-0.357	1.699
Hybrid	1.066	0.706	0.9959	0.509	1.043

5.2. Residual structure and optimization behavior

Residual diagnostics distinguish errors in typical predictions, tail events, and spatially persistent offsets, each of which affects engineering use differently [22]. Figure 4(a) reports primary RMSE values of 5.301, 11.391, 1.804, and 1.066 points for ridge, the branch comparator, the spectral comparator, and the hybrid. The ordering separates the value of spatial information from the additional improvement associated with fused global context. Both spatial neural models improve on the pointwise linear comparator, while the compressed branch representation has the largest error. The shared training budget makes this a comparison of specified implementations: the branch pathway receives spatial means and standard deviations, whereas the spectral pathway receives ordered channel values. Global statistics retain scenario amplitude and variability but collapse the positions of individual events. The branch error is consistent with its restricted spatial representation and characterizes this encoder rather than the full family of DeepONet designs. The hybrid retains full spatial access through its spectral pathway and uses the global coefficients to supply coordinate dependent context. Its lower RMSE can arise from this complementary information structure. Parameter counts and encoder capacities differ across the implementations, so the observed ordering applies to the documented architectural settings. The two point reference drawn across the error panel gives a common diagnostic scale for inspecting the separation between the hybrid and spectral estimates.

Figure 4(b) resolves the signed residual distributions using empirical percentile intervals. Robust regression assessment uses empirical quantiles to describe tail asymmetry while leaving the residual distribution unrestricted [23]. The hybrid median is 0.346 points, with fifth and ninety fifth percentiles of minus 0.590 and 2.283. Its upper tail extends farther from zero, agreeing with the positive pooled bias. The spectral median is minus 0.171, with outer percentiles of minus 3.240 and 2.071, giving a more extended lower tail. Ridge spans minus 6.327 to 8.016 over the same percentile range, and the branch comparator spans minus 19.202 to 14.003. Each interval describes the central 90 percent of evaluated point residuals, while its interquartile box describes the central half. These percentile whiskers measure residual spread rather than confidence in a performance estimate or validation-calibrated predictive coverage. The much wider branch interval indicates that positive and negative deviations persist beyond a small set of extreme cases. For the hybrid, the narrow lower side and more extended positive side suggest a directional amplitude component superimposed on smaller local errors. Tail interpretation also depends on the source population: overlapping event footprints generate the largest targets, and these locations contribute disproportionately to squared error. Separating percentile spread from signed mean error makes it possible to examine this asymmetry directly.

Figure 4(c) compares MAE with absolute bias for each comparator to distinguish residual magnitude from directional cancellation. Hybrid absolute bias is approximately 72 percent of its MAE, compared with approximately 28 percent for the spectral model. The hybrid's smaller MAE therefore coexists with a more pronounced directional component relative to its typical absolute error. These ratios characterize the observed residuals; a formal decomposition of squared error would additionally involve residual variance. For a spatially weighted source-index integral, a persistent offset contributes according to the location weights, while errors of opposing signs can partly cancel. Conversion to a physical segment total requires a length scale and source calibration. Their consequences for maintenance depend on the cost assigned to excessive intervention and to missed high activity segments. The current predictions retain the output of the trained models, preserving the separation between fitting and evaluation. A subsequent calibration study could estimate an offset from validation scenarios and assess the corrected model on an independent evaluation set. Such a design would also allow the residual median and upper tail to be checked after correction. The existing metric pairs suggest that calibration should examine both the global offset and intensity dependent error, because the conditional RMSE rises with source amplitude.

Neural operator benchmarking specifies frequency truncation, channel width, and discretization because these choices affect representational capacity and optimization cost [24]. Figure 4(d) follows the recorded losses for six retained frequency bins, eight spectral channels, and a single mixing block. Hybrid mean squared error falls from 0.02833 at the first recorded epoch to 0.00006448 at epoch 170. The spectral loss falls from 0.05617 to 0.0002605, whereas the branch loss changes from 0.05449 to 0.01306. These mean squared errors are expressed in squared reference-intensity units; conversion to squared index-point units multiplies them by ten thousand. The logarithmic axis reveals the later stages of improvement. The branch training loss approaches a plateau while both spatial models continue to reduce error, consistent with the spatial information compressed by the global summaries. Within the hybrid, the coordinate pathway can describe a smooth spatial modulation while the spectral pathway retains localized forcing structure. This allocation offers a plausible explanation for the fitted loss ordering; isolating its effect requires additional capacity-controlled comparisons. The final recorded loss corresponds to epoch 170 because history is sampled every ten epochs from epoch zero; evaluation uses the completed 180 epoch fit. Runtime entries in the saved report cover model fitting on the execution machine and provide reproducibility context for this compact NumPy implementation. Hardware comparisons would require a controlled timing protocol with matched implementations.

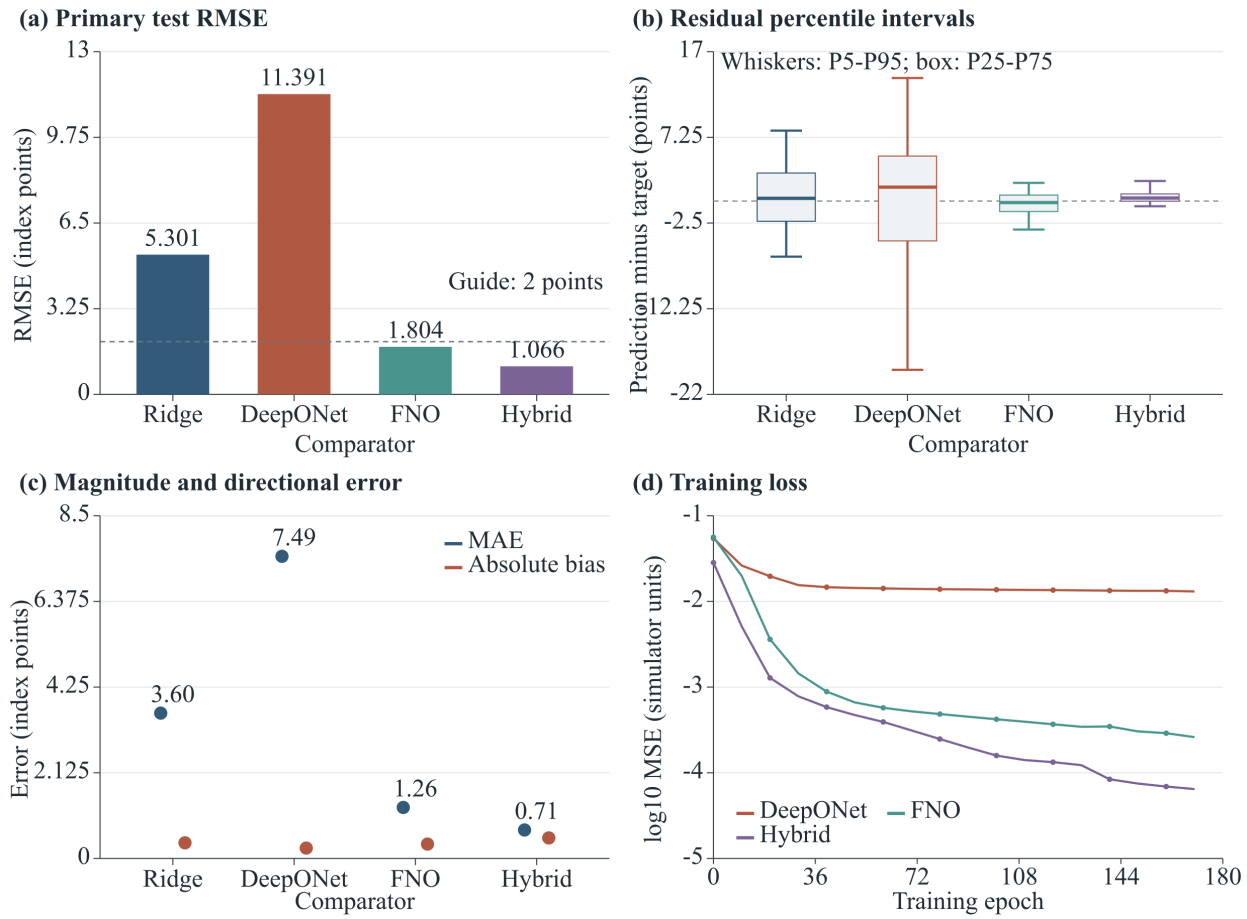


Figure 4. Comparator errors and training diagnostics. (a) Root mean square error bars. (b) Residual percentile boxes and whiskers. (c) Paired markers for mean absolute error and absolute bias. (d) Training mean squared error on a logarithmic scale

5.3. Observation perturbations and transfer limits

Sensor evaluation distinguishes measurement noise from missing observations and calibration drift because each changes the information supplied to an estimator [25]. Figure 5(a) examines six noise standard deviations between zero and 0.12 while preserving the underlying test sources. Hybrid RMSE values are 0.989, 1.007, 1.066, 1.157, 1.272, and 1.552 points; the endpoint increase is 56.89 percent. The spectral comparator rises from 1.588 to 2.980 points, indicating greater sensitivity to particulate perturbation over this range. The branch comparator maintains an RMSE of approximately 11.4 points across the tested noise levels, consistent with the compression of local observation detail into global summaries. Figure 5(b) evaluates random particulate removal followed by interpolation, with hybrid errors of 1.066, 1.059, 1.069, 1.093, 1.160, and 1.169 points as the missing fraction increases from zero to one half. At the largest fraction, the increase over the complete observation setting is 9.69 percent. Interpolation can smooth a portion of the added noise, providing a possible explanation for the small decrease at the first removal level that requires a separate controlled comparison. Trajectory forcing and surface channels remain complete, providing additional model inputs associated with source position and amplitude. Their contribution under this source formulation should be distinguished from statistically independent evidence. The smaller missing-sample response applies to random particulate deletion with these other channels available; uncertain surface measurements or jointly degraded inputs define different tests.

Missing observation methodology separates random deletion from outages associated with environmental state, saturation, and communication failure [26]. Figure 5(c), where the label Wind denotes the transport descriptor, probes channel dependence by permuting each input between test scenarios. Trajectory and concentration permutation raise hybrid RMSE to 14.097 and 8.830 points, respectively, while moisture, loading, speed, and the transport descriptor yield 5.785, 6.708, 5.084, and 1.462. Permutation preserves each channel's spatial structure while disrupting its association with the other channels, thereby measuring dependence within the fitted model. It can create combinations outside the original joint input distribution, so these increases describe predictive dependence rather than causal importance. The larger trajectory response is consistent with the local forcing term in the source formulation, while the concentration response indicates substantial use of the transported observation. Figure 5(d) compares primary and shifted errors after moisture and transport both exceed their training ranges. Hybrid RMSE increases to 1.415 points with a bias of 0.958; the spectral model reaches 1.155 with a bias of 0.365. General regression evaluation under changed inputs considers

target amplitude and the applicability domain alongside absolute prediction error [27]. The shifted source mean falls from 16.60 to 8.01 points and its standard deviation from 16.68 to 7.47. This contraction explains why the branch model's lower shifted RMSE of 5.736 accompanies a smaller R squared of 0.4103. The hybrid and spectral shifted R squared values are 0.9641 and 0.9761, preserving the spectral advantage after normalization by target variability. Because moisture and transport change together, this test does not separate their individual contributions.

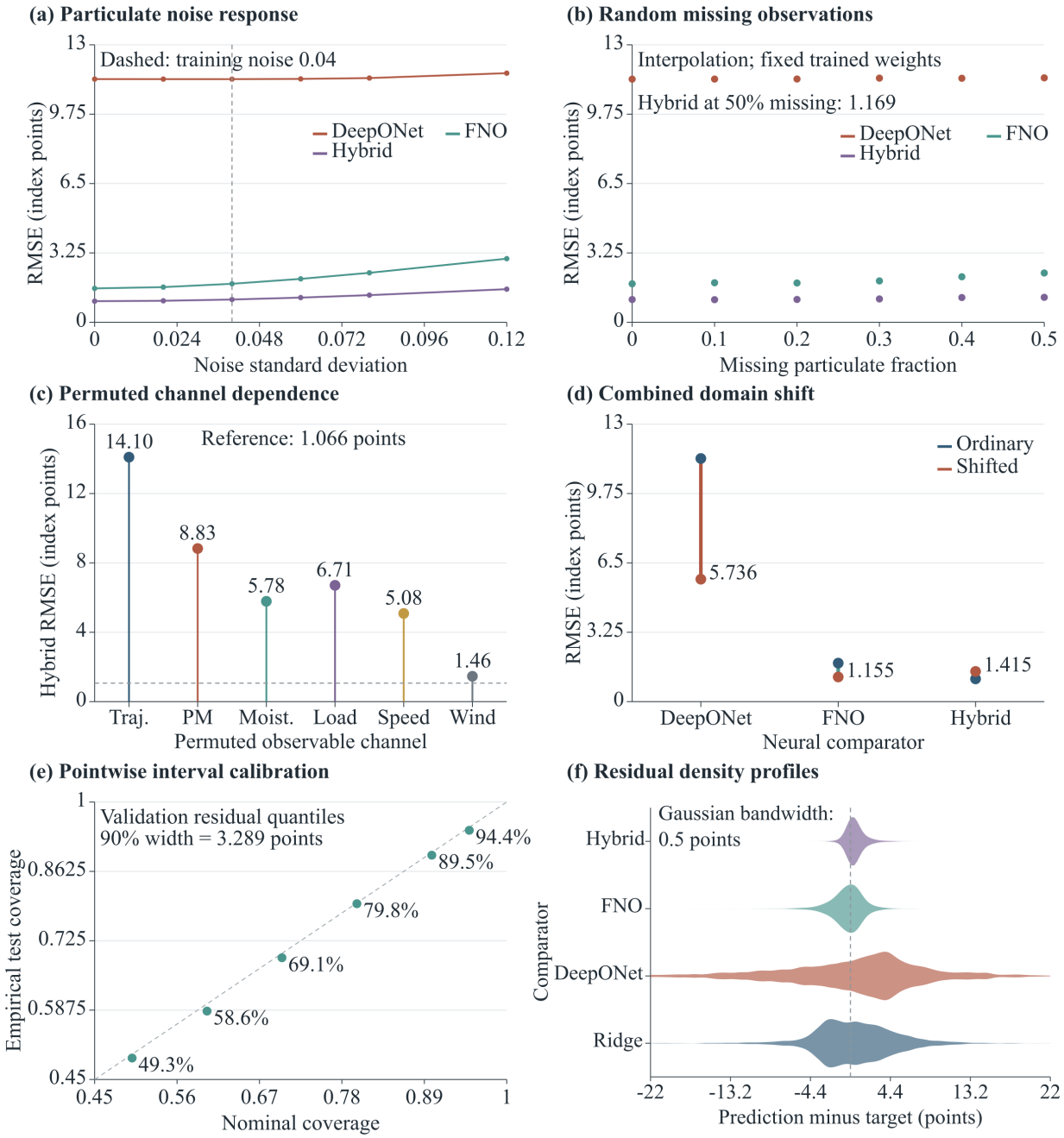


Figure 5. Observation robustness and transfer diagnostics. (a) Noise response curves. (b) Missing observation error markers. (c) Channel permutation lollipops. (d) Primary to shifted error intervals. (e) Pointwise coverage calibration markers. (f) Residual density profiles with equal peak display widths

Prediction interval methodology distinguishes marginal pointwise coverage from coverage conditional on response intensity or environmental state [28]. Figure 5(e) compares nominal levels of 50, 60, 70, 80, 90, and 95 percent with empirical test coverages of 49.25, 58.56, 69.13, 79.84, 89.46, and 94.39 percent. The corresponding full widths are 0.894, 1.145, 1.496, 2.083, 3.289, and 4.447 points. Deviations from nominal coverage remain within 1.5 percentage points for these pooled test residuals, with the 90 percent interval reaching 89.46 percent. The radii are obtained from absolute validation residual quantiles, so asymmetric signed errors are represented by symmetric interval widths. Figure 5(f) displays residual density profiles using a fixed Gaussian visualization bandwidth of 0.5 points. The hybrid profile is concentrated near a small positive error, whereas the spectral profile extends farther toward negative deviations; the branch profile occupies a substantially wider range. Each profile is normalized to the same maximum display width,

making its shape and horizontal spread the relevant comparison rather than total probability mass. Pooled coverage under the primary distribution leaves high-intensity conditional coverage and shifted coverage to separate assessments. Interval calibration and density shape consequently address different properties of the same fitted residual population. Pointwise pooling also combines neighboring locations within scenarios, so uncertainty assessment at the scenario level would require a separate treatment of spatial dependence. The shifted population has its own surface and transport distribution and therefore requires an independent calibration assessment.

Figure 6 examines the dependence of hybrid residuals on target intensity through counts in a 24 by 22 bin display. Marker area increases with count, so the dense low intensity region remains visible alongside the sparsely populated upper tail. The four target bands below 10, from 10 to 20, from 20 to 40, and at least 40 points contain 2531, 1700, 1091, and 438 locations, respectively. These bands summarize conditional errors separately from the two-dimensional display bins. Their RMSE values rise from 0.481 through 0.965 and 1.406 to 2.250 points. The highest band contains 438 of the 5760 evaluated locations, or 7.60 percent, and exhibits an absolute RMSE more than four times that of the lowest band. This absolute-error pattern alone leaves the corresponding relative-error pattern undetermined. The five point diagnostic limits provide a common absolute scale across the response range, while the vertical median marker at 11.55 identifies the division of the target population into equally numerous lower and upper halves. The largest bin contains 1066 locations, emphasizing how strongly pooled residual diagnostics are supported by low intensity predictions. Positive residuals extend farther at some upper range locations, consistent with the hybrid's positive mean bias and upper percentile asymmetry. Because each plotted bin contains spatial point predictions drawn from scenarios with overlapping event footprints, a localized high intensity deviation can affect several adjacent observations. The distinction between concentrated low range support and sparse upper range support is therefore relevant to interpreting both the apparent density and the widening residual envelope.

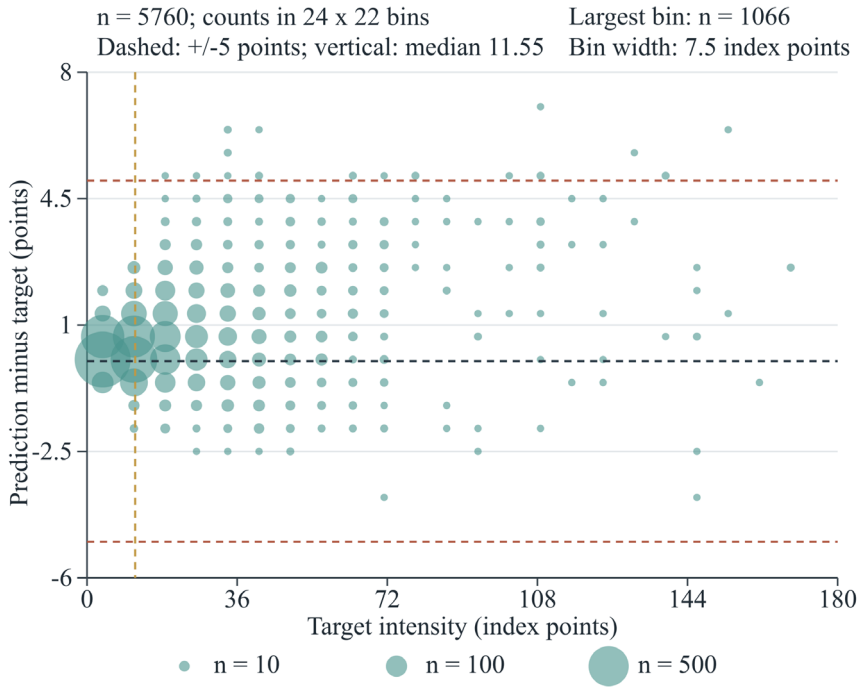


Figure 6. Conditional hybrid residual bubble counts

6. Conclusion

This study evaluated a fixed-basis DeepONet–FNO configuration integrating vehicle trajectory forcing, particulate observations, and road surface descriptors, finding improved intensity-index reconstruction relative to the specified comparators under the primary evaluation conditions.

1. The hybrid model improved reconstruction accuracy. It achieved a test root mean square error of 1.066 index points, a mean absolute error of 0.706 index points, and a coefficient of determination of 0.9959. Compared with the standalone FNO model, root mean square error and mean absolute error decreased by 40.92% and 44.15%, respectively, consistent with complementary information in the evaluated pathways at their documented capacities.

2. The model responded differently to observation noise and missing samples. Increasing the particulate noise standard deviation from zero to 0.12 raised its root mean square error from 0.989 to 1.552 index points. Random removal and interpolation of 50% of particulate samples produced an error of 1.169 index points. These results indicate that

complete trajectory and surface information can provide supplementary constraints when particulate observations are incomplete.

3. The performance advantage depended on environmental conditions. When moisture and transport conditions jointly exceeded their training ranges, the hybrid error increased to 1.415 index points, compared with 1.155 for the FNO comparator, reversing the primary test ranking. This finding identifies distribution change as a material limitation and emphasizes the need to assess input ranges before applying the model to new operating conditions.

Future work should establish physical source calibration, test independent road and seasonal episodes, separate moisture from transport perturbations, and examine structured observation gaps to assess adaptability and conditional reliability.

7. Declarations

7.1 Author Contributions

Ladislav Drápal contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision, project administration, and funding acquisition. Lubomír Hájek contributes to validation, analysis, investigation, data collection, draft preparation. 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] Benabed, A., Fu, E., Arfire, A., & Pousset, P. (2025). Assessment of vehicle-induced road dust resuspension using a novel experimental approach in semi-controlled conditions. Air Quality, Atmosphere & Health, 18(9), 2823–2839. https://doi.org/10.1007/s11869-025-01807-3

[2] Kacer, J., Altmaier, R., Cwiertny, D. M., & O'Shaughnessy, P. T. (2024). Unpaved road particulate matter emission rates and vehicle-induced transient plume characteristics. Environmental Science: Atmospheres, 4(9), 1042–1050. https://doi.org/10.1039/d4ea00055b

[3] Zafra-Mejía, C. A., & Rondón-Quintana, H. A. (2025). Analysis of the potential availability of resuspended heavy metals from urban road sediments: A post-runoff approach. Environmental Pollution, 387, Article 127322. https://doi.org/10.1016/j.envpol.2025.127322

[4] Wagner, S., Funk, C. W., Müller, K., & Raithel, D. J. (2024). The chemical composition and sources of road dust, and of tire and road wear particles—A review. Science of The Total Environment, 926, Article 171694. https://doi.org/10.1016/j.scitotenv.2024.171694

[5] Jiang, W., He, C., Huang, Y., Wang, T., Yuan, D., Wu, W., & Fan, H. (2025). Diffusion of re-suspended dust induced by vehicles: Full-scale simulation and field test. Transportation Research Part D: Transport and Environment, 139, Article 104552. https://doi.org/10.1016/j.trd.2024.104552

[6] Qiu, M., Chen, Z., & Li, X. (2025). A local-to-global progressive association framework for vision-based vehicle trajectory reconstruction. IEEE Transactions on Intelligent Transportation Systems, 26(10), 14927–14941. https://doi.org/10.1109/tits.2025.3583948

[7] Haugen, M., Bühler, P., Schläfle, S., O'Loughlin, D., Saladin, S., Giorio, C., & Boies, A. (2024). Method development and analysis of nanoparticle size fractions from tire-wear emissions. Environmental Science: Atmospheres, 4(9), 1079–1090. https://doi.org/10.1039/d4ea00048j

[8] Patel, M. Y., Vannucci, P. F., Kim, J., Berelson, W. M., & Cohen, R. C. (2024). Towards a hygroscopic growth calibration for low-cost PM2.5 sensors. Atmospheric Measurement Techniques, 17(3), 1051–1060. https://doi.org/10.5194/amt-17-1051-2024

[9] Erraji, H., Franke, P., Lampert, A., Schuldt, T., Tillmann, R., Wahner, A., & Lange, A. C. (2024). The potential of drone observations to improve air quality predictions by 4D-Var. Atmospheric Chemistry and Physics, 24(24), 13913–13934. https://doi.org/10.5194/acp-24-13913-2024

[10] Al Aawar, E., Resifi, S., Jebari, H., & Hoteit, I. (2025). Bayesian source identification with dual hierarchical neural networks for urban air pollution. Journal of Advances in Modeling Earth Systems, 17(4), Article e2024MS004790. https://doi.org/10.1029/2024ms004790

[11] Veratti, G., Bigi, A., Stortini, M., Teggi, S., & Ghermandi, G. (2024). Measurement report: Source attribution and estimation of black carbon levels in an urban hotspot of the central Po Valley – an integrated approach combining high-resolution dispersion modelling and micro-aethalometers. Atmospheric Chemistry and Physics, 24(18), 10475–10512. https://doi.org/10.5194/acp-24-10475-2024

[12] Kaminski, M. D., Daiyega, N., & Magnuson, M. (2024). A review of the resuspension of radioactively contaminated particles by vehicle and pedestrian traffic—Current theory, practice, gaps, and needs. Health Physics, 126(4), 216–240. https://doi.org/10.1097/hp.0000000000001797

[13] Park, S.-J., Lugon, L., Jacquot, O., Kim, Y., Baudic, A., D'Anna, B., Di Antonio, L., Di Biagio, C., Dugay, F., Favez, O., Ghersi, V., Gratien, A., Kammer, J., Petit, J.-E., Sanchez, O., Valari, M., Vigneron, J., & Sartelet, K. (2025). Population exposure to outdoor NO2, black carbon, and ultrafine and fine particles over Paris with multi-scale modelling down to the street scale. Atmospheric Chemistry and Physics, 25(6), 3363–3387. https://doi.org/10.5194/acp-25-3363-2025

[14] Mathieu-Campbell, M. E., Guo, C., Grieshop, A. P., & Richmond-Bryant, J. (2024). Calibration of PurpleAir low-cost particulate matter sensors: Model development for air quality under high relative humidity conditions. Atmospheric Measurement Techniques, 17(22), 6735–6749. https://doi.org/10.5194/amt-17-6735-2024

[15] Liu, S., Yu, Y., Zhang, T., Liu, H., Liu, X., & Meng, D. (2025). Architectures, variants, and performance of neural operators: A comparative review. Neurocomputing, 648, Article 130518. https://doi.org/10.1016/j.neucom.2025.130518

[16] Azizzadenesheli, K., Kovachki, N., Li, Z., Liu-Schiaffini, M., Kossaifi, J., & Anandkumar, A. (2024). Neural operators for accelerating scientific simulations and design. Nature Reviews Physics, 6(5), 320–328. https://doi.org/10.1038/s42254-024-00712-5

[17] Auras, A., Gandikota, K. V., Droege, H., & Moeller, M. (2024). Robustness and exploration of variational and machine learning approaches to inverse problems: An overview. GAMM-Mitteilungen, 47(4), Article e202470003. https://doi.org/10.1002/gamm.202470003

[18] Singh, G., Moncrieff, G., Venter, Z., Cawse-Nicholson, K., Slingsby, J., & Robinson, T. B. (2024). Uncertainty quantification for probabilistic machine learning in earth observation using conformal prediction. Scientific Reports, 14(1), Article 16166. https://doi.org/10.1038/s41598-024-65954-w

[19] John, K., Saurette, D. D., & Heung, B. (2025). The problematic case of data leakage: A case for leave-profile-out cross-validation in 3-dimensional digital soil mapping. Geoderma, 455, Article 117223. https://doi.org/10.1016/j.geoderma.2025.117223

[20] Bataka, E. P., Maletsika, P., & Nakas, C. T. (2024). Formal assessment of agreement and similarity between an open-source and a reference industrial device with an application to a low-cost pH logger. Sensors, 24(2), Article 490. https://doi.org/10.3390/s24020490

[21] Linnenbrink, J., Milà, C., Ludwig, M., & Meyer, H. (2024). kNNDM CV: k-fold nearest-neighbour distance matching cross-validation for map accuracy estimation. Geoscientific Model Development, 17(15), 5897–5912. https://doi.org/10.5194/gmd-17-5897-2024

[22] Kaniuka, J., Ostrysz, J., Groszyk, M., Bieniek, K., Cyperski, S., & Domański, P. D. (2024). Multicriteria machine learning model assessment—Residuum analysis review. Electronics, 13(5), Article 810. https://doi.org/10.3390/electronics13050810

[23] Tyralis, H., & Papacharalampous, G. (2024). A review of predictive uncertainty estimation with machine learning. Artificial Intelligence Review, 57(4), Article 94. https://doi.org/10.1007/s10462-023-10698-8

[24] Cao, Q., Goswami, S., & Karniadakis, G. E. (2024). Laplace neural operator for solving differential equations. Nature Machine Intelligence, 6(6), 631–640. https://doi.org/10.1038/s42256-024-00844-4

[25] Barkjohn, K. K., Clements, A., Mocka, C., Barrette, C., Bittner, A., Champion, W., Gantt, B., Good, E., Holder, A., Hillis, B., Landis, M. S., Kumar, M., MacDonald, M., Thoma, E., Dye, T., Archer, J.-M., Bergin, M., Mui, W., Feenstra, B., … Zimmerman, N. (2024). Air quality sensor experts convene: Current quality assurance considerations for credible data. ACS ES&T Air, 1(10), 1203–1214. https://doi.org/10.1021/acsestair.4c00125

[26] Decorte, T., Mortier, S., Lembrechts, J. J., Meysman, F. J. R., Latré, S., Mannens, E., & Verdonck, T. (2024). Missing value imputation of wireless sensor data for environmental monitoring. Sensors, 24(8), Article 2416. https://doi.org/10.3390/s24082416

[27] Dutschmann, T.-M., Schlenker, V., & Baumann, K. (2024). Chemoinformatic regression methods and their applicability domain. Molecular Informatics, 43(7), Article e202400018. https://doi.org/10.1002/minf.202400018

[28] 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