FRAME: RETHINKING SPATIO-TEMPORAL FORECASTING FROM OBSERVED BOUNDARIES TO FUTURE FIELDS

Anonymous authors
Paper under double-blind review

ABSTRACT

Spatio-temporal forecasting asks how an observed relational field should extend into an unknown future. We cast this task as boundary-conditioned extension: history supplies a measured boundary, directed relations define its geometry, and the graph–horizon product domain supports the future field. We introduce *FRAME* (Future-field Reconstruction through Action-Minimizing Extension), which separates this construction into coordinate-consistent boundary representation, explicit continuation of measured structure, and joint inference of innovation. Robust local affine fibers and sample-conditioned whitened Galerkin projections coordinate the boundary, while a constant-preserving affine trace carries observed level and trajectory forward. Innovation is defined by the unique minimizer of a strongly convex robust action coupling conservative graph exchange, temporal variation, curvature, and boundary attachment, and is approximated by learned solver steps. Our analysis establishes affine factorization, non-expansive projection, uniqueness, and bounded conservative structural influence. Direct numerical verification further shows that every learned update descends the stated action toward its variational target and that four steps closely reproduce the prediction induced by the converged target. Across twelve tasks spanning ten datasets, including ChinaTemp and ChinaWind, FRAME attains the lowest MAE on every task and the lowest RMSE on all six highway and both meteorological benchmarks.

1 INTRODUCTION

Many dynamical systems are observed as fields over related entities: environmental sensors, power-network components, industrial units, communication links, mobility regions, or spatial cells. Their prediction target is not merely a collection of independent numbers but a future field whose relational variation, horizon-wise evolution, and attachment to observation should agree. We therefore ask a more fundamental question than how to build another forecasting architecture: *what defines a valid extension of an observed spatio-temporal field?* In our formulation, observed history is a measured boundary and the original directed graph $\mathcal{G}$ specifies its relational geometry. The induced conservative support $\mathcal{G}_c$, together with the requested horizon path P_H, defines the unknown extension domain. Forecasting becomes the construction of a future field on $\mathcal{G}_c \square P_H$.

Prior spatio-temporal forecasters coordinate dependence primarily inside learned representations; Section 2 reviews the corresponding architectural, dynamical, and optimization-inspired views. FRAME instead makes structured future innovation an explicit boundary-conditioned variational target, so graph exchange, temporal variation, curvature, and boundary attachment act directly on the reduced future field. The question shifts from choosing a dependence encoder to defining the extension that the future field should satisfy. This yields three dependent requirements. **Challenge I: how can heterogeneous node histories define a meaningful relational boundary?** Raw levels and units should not obscure comparable dynamics, while sample-dependent global coordinates must remain geometrically controlled. **Challenge II: given that boundary, what should be continued rather than regenerated?** The observed level and local trajectory contain inherited structure that should have an explicit path into the future. **Challenge III: after continuation is separated, how should innovation be determined jointly?** Graph disagreement, temporal variation, curvature,

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

and boundary attachment must act on one future field rather than remain implicit in a decoded representation.

FRAME answers this chain through three corresponding principles. **Boundary:** robust affine fibers and whitened sample-conditioned Galerkin bases construct a coordinate-consistent observed state. **Continuation:** a constant-preserving affine trace carries inherited level and local trajectory forward. **Innovation:** a robust action treats graph and horizon differences as derivatives of one future field on $\mathcal{G}_c \square P_H$ and attaches its first slice to observation. For fixed conditioning, the strongly convex action defines a unique coordinated extension. Its $\log\cosh$ potentials bound local influence, and antisymmetric graph flux redistributes state without graph-induced net drift. FRAME approximates this target through learned preconditioned steps and combines affine continuation, action correction, and full-width context in the final readout.

Our contributions are:

- We formulate forecasting as boundary-conditioned variational extension on $\mathcal{G}_c \square P_H$: a reduced future innovation field is jointly governed by relational exchange, temporal variation, curvature, and boundary attachment.
- We construct a coordinate-consistent boundary from robust node-local affine fibers and sample-conditioned whitened Galerkin coordination, separating physical scale from comparable dynamics.
- We separate inherited structure from innovation: a constant-preserving affine trace carries observation-supported structure, while a strongly convex action jointly determines innovation over graph and horizon geometry.
- We establish affine factorization, projection stability, uniqueness, and bounded conservative influence, verify the finite solver against the numerical action target, and demonstrate the formulation on twelve highway, urban-flow, and meteorological forecasting tasks.

2 Related Work

Relational spatio-temporal forecasting. Diffusion recurrence embeds directed transport in sequential evolution (Li et al., 2018), whereas graph–temporal convolution alternates spatial and temporal filtering (Yu et al., 2018). Adaptive graph filters learn node-specific relations (Bai et al., 2020); attention models capture propagation delay (Jiang et al., 2023) or global adaptive embeddings (Liu et al., 2023). FRAME instead uses relations to coordinate the measured boundary and conservative exchange within the predicted field.

Dynamical and operator views. Continuous-time models express latent graph evolution through neural differential equations (Fang et al., 2021) or controlled dynamics (Choi et al., 2022). Neural operators learn field mappings (Kovachki et al., 2023), with Fourier kernels for global interaction (Li et al., 2021) and Galerkin projections for learned trial spaces (Cao, 2021). FRAME uses operator geometry to construct the boundary and drive of an explicit extension problem.

Variational and optimization-inspired prediction. Algorithm unrolling turns iterative inference into trainable finite-depth networks (Monga et al., 2021), while differentiable optimization embeds structured programs in end-to-end models (Amos & Kolter, 2017). FRAME makes the reduced future innovation field on $\mathcal{G}_c \square P_H$ the variable of a sample-conditioned action: representations construct the boundary, the affine trace preserves measured structure, and strong convexity defines a unique coordinated target approximated by a finite solver.

3 Method

FRAME realizes boundary-conditioned future extension through three principles: a coordinate-consistent relational boundary, an explicit constant-preserving continuation, and action-coordinated innovation on the conservative graph–horizon product lattice. Figure 1 summarizes this computation before the individual components are defined. The final prediction combines these structured components with full-width context before returning to the node-local physical coordinate.

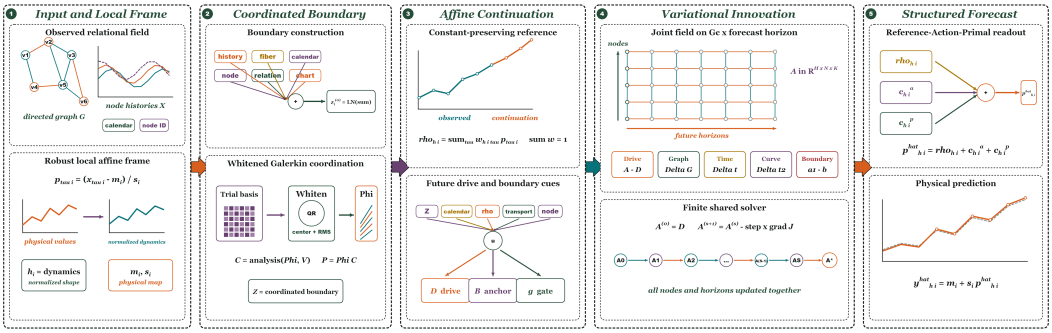

Figure 1: FRAME transforms an observed relational boundary into a coordinated future field.

3.1 Problem Formulation and Notation

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a directed relational graph with $N = |\mathcal{V}|$ nodes. It induces the conservative support graph $\mathcal{G}_c = (\mathcal{V}, \mathcal{E}_c)$, where $\{i, j\} \in \mathcal{E}_c$ whenever $(i, j) \in \mathcal{E}$ or $(j, i) \in \mathcal{E}$. For one sample, the model receives

$$\mathbf{X} \in \mathbb{R}^{T \times N \times C_{\text{in}}}, \qquad \widehat{\mathbf{Y}} = \mathcal{F}_\theta(\mathbf{X}, \mathcal{G}) \in \mathbb{R}^{H \times N}, \tag{1}$$

where T and H are the observed and forecast lengths. We write $x_{\tau i} = X_{\tau,i,1}$ for the scalar forecasting variable at observed step $\tau \in \{1, \dots, T\}$ and node $i \in \{1, \dots, N\}$. The remaining input channels provide future-extendable calendar indices, namely time of day and day of week. The batch dimension is omitted throughout. We use D for the boundary-state width, r for the Galerkin rank, K for the reduced action width, and S for the number of solver steps. The observed and forecast lengths T and H may differ.

The signal channel is standardized using statistics from the training split. FRAME additionally creates a sample-wise local affine coordinate at each node; its final output is returned through this local coordinate and then through the training-set inverse transform. All hidden maps below act on the last tensor dimension unless stated otherwise.

3.2 Relational Boundary Construction

Directed relational coordinates. A meaningful boundary must encode where a node lies in directed relational geometry, including asymmetric source- and target-side roles. Local aggregation alone does not provide a stable signed global chart. FRAME therefore constructs a fixed coordinate matrix $R_G = \mathrm{RelCoord}(G) \in \mathbb{R}^{N \times r}$. The implemented map converts directed edge lengths to a two-hop affinity, applies a truncated singular decomposition, combines its source- and target-side directions, and normalizes every coordinate over nodes. Consequently, rows of R_G encode directed relational position through signed source- and target-side coordinates. FRAME deliberately assigns the two relational geometries different roles: directionality locates the observed boundary, whereas the antisymmetric pairwise flux on $\mathcal{G}_c$ coordinates future innovation without introducing graph-induced net drift.

Robust affine boundary fiber. Raw node histories entangle dynamical shape with local offset and amplitude. A common future operator should expose comparable dynamics without discarding the coordinate needed to reconstruct physical values. For node i, FRAME therefore factors the observed trajectory into a robust local affine coordinate:

$$\begin{aligned} m_i &= \mathrm{median}_\tau\, x_{\tau i}, \qquad s_i = \max\left(\epsilon, \frac{1}{T}\sum_{\tau=1}^{T} |x_{\tau i} - m_i|\right), \\ p_{\tau i} &= \frac{x_{\tau i} - m_i}{s_i}, \qquad \delta_{\tau i} = p_{\tau i} - p_{\tau-1,i}, \qquad \kappa_{\tau i} = \delta_{\tau i} - \delta_{\tau-1,i}, \\ h_i &= [p_i; \delta_i; \kappa_i] \in \mathbb{R}^{3T}, \qquad b_i = [m_i, \log s_i, \delta_{Ti}, \kappa_{Ti}]^\top \in \mathbb{R}^4. \end{aligned} \tag{2}$$

Here $p_i = (p_{\tau i})_{\tau=1}^{T}$, $\delta_i = (\delta_{\tau i})_{\tau=1}^{T}$, and $\kappa_i = (\kappa_{\tau i})_{\tau=1}^{T}$. The initial differences are set to zero and $\epsilon > 0$ ensures a positive scale. The pair (m_i, s_i) defines the local physical coordinate, while the terminal slope and curvature summarize the dynamical state at the observed boundary.

Let $E_{\mathrm{hist}} : \mathbb{R}^{3T} \to \mathbb{R}^D$ and $E_{\mathrm{fib}} : \mathbb{R}^4 \to \mathbb{R}^D$ be learned lifts. Let $e_T^{\mathrm{cal}} \in \mathbb{R}^D$ denote the sum of time-of-day and day-of-week embeddings at the final observation; let $e_i^{\mathrm{node}}, e_i^{\mathrm{chart}} \in \mathbb{R}^D$ be learned node and boundary-chart embeddings; and let $E_G : \mathbb{R}^r \to \mathbb{R}^D$ lift the i-th row r_i of R_G. The initial boundary state is

$$z_i^{(0)} = \mathrm{LN}_{\mathrm{in}}\left(E_{\mathrm{hist}}(h_i) + E_{\mathrm{fib}}(b_i) + e_T^{\mathrm{cal}} + e_i^{\mathrm{node}} + E_G(r_i) + e_i^{\mathrm{chart}}\right) \in \mathbb{R}^D. \tag{3}$$

Stacking all nodes gives $Z^{(0)} \in \mathbb{R}^{N \times D}$.

3.3 Sample-Conditioned Whitened Coordination

The boundary chart should adapt to the current sample, but an unconstrained sample-dependent basis has arbitrary scale and correlation, making its projection geometry ill conditioned. FRAME resolves this ambiguity by whitening every learned trial basis. For block $\ell \in \{0, \dots, L-1\}$, let $\bar{Z}^{(\ell)} = \mathrm{LN}_1^{(\ell)}(Z^{(\ell)})$. Its unwhitened node basis is

$$\widetilde{\Phi}^{(\ell)} = \bar{Z}^{(\ell)} W_\phi^{(\ell)} + \gamma_G^{(\ell)} R_G M_G^{(\ell)} \in \mathbb{R}^{N \times r}, \tag{4}$$

where $W_\phi^{(\ell)} \in \mathbb{R}^{D \times r}$, $M_G^{(\ell)} \in \mathbb{R}^{r \times r}$, and $\gamma_G^{(\ell)} > 0$ are learned. Let $\mathcal{N}(\cdot)$ denote column centering followed by root-mean-square normalization over nodes. A reduced QR factorization gives

$$\Phi^{(\ell)} = \sqrt{N}\ \mathrm{qf}\left(\mathcal{N}\left(\widetilde{\Phi}^{(\ell)}\right)\right), \qquad \frac{1}{N}\left(\Phi^{(\ell)}\right)^\top \Phi^{(\ell)} = I_r. \tag{5}$$

Thus, the basis remains sample conditioned while its empirical Gram geometry is controlled.

With $V^{(\ell)} = \bar{Z}^{(\ell)} W_v^{(\ell)} \in \mathbb{R}^{N \times D}$, Galerkin analysis and synthesis are

$$C^{(\ell)} = \frac{1}{N}\left(\Phi^{(\ell)}\right)^\top V^{(\ell)} \in \mathbb{R}^{r \times D}, \qquad P^{(\ell)} = \Phi^{(\ell)} C^{(\ell)} \in \mathbb{R}^{N \times D}. \tag{6}$$

The synthesized projection is injected through a learned gated residual connection and followed by a standard pre-normalized position-wise MLP update. After L blocks, $Z = \mathrm{LN}_{\mathrm{bd}}(Z^{(L)}) \in \mathbb{R}^{N \times D}$ is the globally coordinated boundary, with row $z_i^\top$.

3.4 Boundary-Conditioned Affine Trace

Not every future component is new behavior. If inherited level and local trajectory are left implicit, the latent solver must regenerate structure already measured at the boundary. FRAME instead separates a sample- and node-conditioned affine reference, a horizon-dependent drive, and a reduced boundary anchor. The action solver can then determine innovation around an explicit continuation path.

For horizon h, a learned base row $\pi_h = \mathrm{softmax}(\ell_h) \in \mathbb{R}^T$ provides an initial history weighting. The coordinated boundary produces a horizon-specific query $q_{hi} = \mathrm{reshape}_h(E_q(z_i)) \in \mathbb{R}^{d_q}$, which defines adaptation scores $r_{hi\tau} = q_{hi}^\top k_\tau / \sqrt{d_q} + t_{h\tau}$, where $k_\tau \in \mathbb{R}^{d_q}$ and $t_{h\tau} \in \mathbb{R}$ are learned history keys and horizon–history tilts.

Let $\mathcal{A} = \{w \in \mathbb{R}^T : \mathbf{1}^\top w = 1\}$, and let $\mathrm{Proj}_{\mathcal{A}}$ subtract the mean excess from a vector. The affine trace weights and reference are

$$w_{hi} = \mathrm{Proj}_{\mathcal{A}}\left(\pi_h + \gamma_{\mathrm{tr}} \tanh(r_{hi})\right), \qquad \rho_{hi} = \sum_{\tau=1}^{T} w_{hi\tau} p_{\tau i}. \tag{7}$$

By construction, $\sum_{\tau=1}^{T} w_{hi\tau} = 1$, where $\gamma_{\mathrm{tr}} > 0$ is a fixed bounded adaptation amplitude. Because the weights lie on an affine hyperplane rather than a probability simplex, the trace can express both local continuation and controlled extrapolation while remaining in the observed node-specific coordinate.

The terminal slope and curvature provide complementary directional information at the observation boundary. FRAME encodes this cue through

$$\begin{aligned} \xi_{hi} &= h\delta_{Ti} + \frac{1}{2}h(h-1)\kappa_{Ti}, \\ k_{hi}^{\mathrm{tr}} &= \gamma_{\mathrm{fib}} \tanh\big(E_{\mathrm{tr}}\left([3\tanh(\xi_{hi}/3), \tanh(\delta_{Ti} + h\kappa_{Ti}), \kappa_{Ti}]^{\top}\right)\big) \in \mathbb{R}^{D}. \end{aligned} \tag{8}$$

This transport supplies a bounded directional boundary cue rather than an independent hand-crafted forecast.

Let $e_h^{\mathrm{fut}} \in \mathbb{R}^D$ combine the future time-of-day, day-of-week, and horizon embeddings. With $E_\rho : \mathbb{R} \to \mathbb{R}^D$, FRAME constructs

$$\begin{aligned} u_{hi} &= z_i + e_h^{\mathrm{fut}} + e_i^{\mathrm{node}} + E_\rho(\rho_{hi}) + k_{hi}^{\mathrm{tr}} \in \mathbb{R}^D, \\ d_{hi} &= D_\theta\left(\mathrm{LN}_{\mathrm{drv}}(u_{hi})\right) \in \mathbb{R}^K, \\ b_i^a &= E_b(z_i) \in \mathbb{R}^K, \qquad g_i = \frac{1}{2}\mathbf{1} + \sigma(E_g(z_i)) \in (0.5, 1.5)^K. \end{aligned} \tag{9}$$

Here d_{hi} is the horizon-dependent unconstrained future drive, b_i^a is the reduced boundary anchor, and g_i modulates graph exchange. Thus, ρ_{hi} determines the explicit continuation path, whereas d_{hi} and b_i^a provide the drive and boundary condition for variational coordination.

3.5 Variational Action on the Conservative Graph–Horizon Lattice

After continuation is separated through ρ_{hi}, the remaining innovation must be coherent on both axes of the future domain. An unconstrained drive supplies evidence but does not assign graph exchange, horizon variation, curvature, and boundary attachment explicit structural roles. FRAME therefore treats the reduced future innovation field as the variable of a single robust action over $\mathcal{G}_c \square P_H$, directly defining the structured innovation consumed by the readout. Let $A = [a_{hi}] \in \mathbb{R}^{H\times N\times K}$ denote the reduced action field and $D = [d_{hi}]$ the drive of the same shape. All H future states are present simultaneously: the drive provides an unconstrained proposal, while the action coordinates it according to graph consistency, temporal continuity, temporal curvature, and attachment to the observed boundary.

Each conservative pair $\{i, j\} \in \mathcal{E}_c$ has a fixed positive degree-normalized weight ω_{ij}, and its boundary-conditioned gate is $g_{ij} = \frac{1}{2}(g_i + g_j) \in \mathbb{R}^K$. For positive vectors $\kappa, \beta \in \mathbb{R}^K_+$, define

$$\phi_{\kappa,\beta}(v) = \sum_{k=1}^{K} \frac{\kappa_k}{\beta_k} \log\cosh(\beta_k v_k), \qquad v \in \mathbb{R}^K. \tag{10}$$

Its derivative is bounded component-wise by κ and behaves quadratically near zero, encouraging smooth agreement for small discrepancies while limiting the influence of large graph or temporal jumps.

Let $\Delta_G a_{h,ij} = a_{hj} - a_{hi}$ for $\{i, j\} \in \mathcal{E}_c$, $\Delta_t a_{hi} = a_{h+1,i} - a_{hi}$, and $\Delta_t^2 a_{hi} = a_{h+2,i} - 2a_{h+1,i} + a_{hi}$. With positive learned conductivities $\kappa_G, \kappa_t, \kappa_c \in \mathbb{R}^K_+$, slopes $\beta_G, \beta_t, \beta_c, \beta_b \in \mathbb{R}^K_+$, and boundary strength $\lambda_b > 0$, the learned action is

$$
\begin{aligned}
\mathcal{J}(A \mid D, B) = \frac{1}{2} \sum_{h=1}^{H} \sum_{i=1}^{N} \|a_{hi} - d_{hi}\|_2^2 \\
+ \sum_{h=1}^{H} \sum_{\{i,j\} \in \mathcal{E}_c} \omega_{ij} \phi_{g_{ij} \odot \kappa_G, \beta_G}(\Delta_G a_{h,ij}) \\
+ \sum_{i=1}^{N} \sum_{h=1}^{H-1} \phi_{\kappa_t, \beta_t}(\Delta_t a_{hi}) \\
+ \sum_{i=1}^{N} \sum_{h=1}^{H-2} \phi_{\kappa_c, \beta_c}(\Delta_t^2 a_{hi}) \\
+ \sum_{i=1}^{N} \phi_{\lambda_b \mathbf{1}, \beta_b}(a_{1i} - b_i^a),
\end{aligned}
\tag{11}
$$

where $B = [b_i^a]$. The five terms respectively represent fidelity to the learned drive, conservative graph coordination, first-order temporal consistency, temporal-curvature control, and attachment of the first future state to the observed boundary. All five terms act on the same unknown field A, so their interaction defines one joint innovation problem over graph and horizon geometry. Sums over empty horizon ranges vanish: when $H = 1$, both temporal terms disappear and the same formulation reduces to drive fidelity, conservative graph exchange, and boundary attachment. For multi-step horizons, the first-difference null space contains constant trajectories, whereas the second-difference null space contains affine trajectories; curvature therefore controls changes of trend without adding curvature cost to persistent linear motion.

Differentiating these robust potentials yields bounded graph, temporal, curvature, and boundary fluxes. Let F^G, F^t, F^c, F^b denote the corresponding flux fields, and let $\Delta_G^\top$, $\Delta_t^\top$, and $(\Delta_t^2)^\top$ denote the adjoint scatter operators. The action gradient can then be written compactly as

$$
\nabla_A \mathcal{J} = A - D + \Delta_G^\top F^G + \Delta_t^\top F^t + (\Delta_t^2)^\top F^c + e_1 F^b. \tag{12}
$$

For each graph pair, the same graph flux is subtracted from one endpoint and added to the other, so the graph contribution sums to zero over nodes and preserves pairwise conservative exchange.

FRAME initializes the action field at the learned drive, $A^{(0)} = D$, and performs S shared preconditioned iterations:

$$
A^{(s+1)} = A^{(s)} - \eta \Pi \odot \nabla_A \mathcal{J}\left(A^{(s)} \mid D, B\right), \qquad s = 0, \ldots, S-1, \tag{13}
$$

where $\eta > 0$ is a learned bounded step and $\Pi \in \mathbb{R}_+^K$ is a learned diagonal preconditioner broadcast over horizons and nodes. Every update acts on all H horizons simultaneously, allowing graph, temporal, curvature, and boundary terms to coordinate the complete reduced innovation field jointly.

3.6 Reference–Action–Primal Readout

FRAME assigns distinct roles to its three readout paths: the affine continuation preserves inherited structure, the action branch determines explicitly structured innovation, and the full-width context supplies complementary predictive information. The normalized forecast is reconstructed through

$$
\hat{p}_{hi} = \underbrace{\rho_{hi}}_{\text{affine continuation}} + \underbrace{c_{hi}^a}_{\text{variational correction}} + \underbrace{c_{hi}^p}_{\text{context correction}}. \tag{14}
$$

with

$$
c_{hi}^a = \lambda_a R_\theta \left(\mathrm{LN}_{\mathrm{out}}\left([a_{hi}; d_{hi}; a_{hi} - b_i^a]\right)\right), \qquad c_{hi}^p = \lambda_p P_\theta(u_{hi}), \tag{15}
$$

where $a_{hi} = A_{hi}^{(S)}$, $R_\theta : \mathbb{R}^{3K} \to \mathbb{R}$, and $P_\theta : \mathbb{R}^D \to \mathbb{R}$. The action correction reads the jointly regularized future state together with its drive and boundary displacement, while the parallel context correction carries complementary predictive information from the full D-dimensional future representation.

Finally, the normalized prediction is returned through the node-specific affine fiber:

$$\hat{y}_{hi} = m_i + s_i \hat{p}_{hi}. \tag{16}$$

The training-set inverse transform restores physical units. All parameters are optimized end to end using physical-scale masked MAE under the official validity mask.

3.7 Theoretical Analysis

The analysis follows the boundary–continuation–innovation decomposition: affine factorization and whitening make the boundary well posed, the affine trace preserves inherited structure, and the action defines a unique future extension with controlled structural influence. Complete proofs and the product-lattice operator form are given in Appendix A.

Proposition 1 (Local affine factorization)**.** *For a node-wise change $x'_{\tau i} = a_i x_{\tau i} + b_i$, $a_i > 0$, with the scale floor inactive, $m'_i = a_i m_i + b_i$, $s'_i = a_i s_i$, and $(p'_i, \delta'_i, \kappa'_i) = (p_i, \delta_i, \kappa_i)$. Hence the dynamic channels are coordinate invariant while (m_i, s_i) retains the physical map.*

Proposition 2 (Stable whitened projection)**.** *If the normalized trial basis has full column rank, $\mathcal{P}_\Phi = N^{-1}\Phi\Phi^\top$ is an orthogonal projector: $\mathcal{P}_\Phi^\top = \mathcal{P}_\Phi$, $\mathcal{P}_\Phi^2 = \mathcal{P}_\Phi$, and $\|\mathcal{P}_\Phi V\|_F \le \|V\|_F$.*

Because $\mathbf{1}^\top w_{hi} = 1$, the affine trace preserves constants and is translation consistent: $\rho_{hi}(p_i + c\mathbf{1}) = \rho_{hi}(p_i) + c$. Innovation is therefore assigned to the coordinated action rather than mixed with inherited level.

Theorem 1 (Unique coordinated variational extension)**.** *For fixed D, B, boundary-conditioned gates, and positive action parameters, $\mathcal{J}(A \mid D, B)$ is coercive and 1-strongly convex in A. It therefore has the unique minimizer $A^\star = \arg\min_A \mathcal{J}(A \mid D, B)$, approximated by the learned finite-step solver $A^{(S)}$.*

Proposition 3 (Bounded conservative structural flux)**.** *Since $\nabla_v \phi_{\kappa,\beta}(v) = \kappa \odot \tanh(\beta \odot v)$, every local flux is bounded by its conductivity; moreover $\sum_i [\Delta_G^\top F^G]_{hi} = 0$ for every horizon and channel. Thus graph coordination redistributes state conservatively while saturating large discrepancies.*

The finite solver is also checked directly against this variational target. Appendix B.3 evaluates every learned iterate under the exact action, compares it with a high-accuracy numerical minimizer, and verifies progressive reduction of both action suboptimality and first-order residual.

4 Experiments

4.1 Datasets

We evaluate FRAME on PeMS03, PeMS04, PeMS07, and PeMS08 (Song et al., 2020), PeMSD7(M) and PeMSD7(L) (Yu et al., 2018), and the T-Drive and CHIBike urban-flow benchmarks (Hu et al., 2026). Table 1 summarizes their domains and forecasting protocols. The six highway datasets use twelve observed steps to predict twelve future steps. T-Drive and CHIBike use six observations to predict the next inflow or outflow field. All datasets retain their established chronological splits, graph constructions, sampling intervals, and evaluation masks.

Table 1: Dataset statistics and forecasting protocols.

Dataset	Domain	#Nodes	#Edges	Interval	Protocol
PeMS03	Traffic Flow	358	547	5 min	$12 \to 12$
PeMS04	Traffic Flow	307	340	5 min	$12 \to 12$
PeMS07	Traffic Flow	883	866	5 min	$12 \to 12$
PeMS08	Traffic Flow	170	303	5 min	$12 \to 12$
PeMSD7(M)	Traffic Speed	228	1,132	5 min	$12 \to 12$
PeMSD7(L)	Traffic Speed	1,026	10,150	5 min	$12 \to 12$
T-Drive	Urban Flow	1,024	7,812	60 min	$6 \to 1$
CHIBike	Urban Flow	270	1,966	30 min	$6 \to 1$

4.2 BASELINES

The baseline set is ARIMA (Box et al., 2015); FC-LSTM and GRU-ED (Hochreiter & Schmidhuber, 1997; Cho et al., 2014); DCRNN (Li et al., 2018); ASTGCN and MSTGCN (Guo et al., 2019); STGCN (Yu et al., 2018); STGODE (Fang et al., 2021); STSGCN (Song et al., 2020); TFormer (Yan et al., 2022); PDFormer (Jiang et al., 2023); STAEformer (Liu et al., 2023); STGNCDE (Choi et al., 2022); CaST (Xia et al., 2023); RAST (Ruan et al., 2026); and MFormer (Hu et al., 2026).

4.3 EXPERIMENTAL SETTINGS

All experiments are conducted on an NVIDIA GeForce RTX 4090 GPU with PyTorch and automatic mixed precision. Optimization uses AdamW, learning-rate warm-up followed by cosine decay, gradient clipping, exponential moving-average weights, and validation-based early stopping. Normalization statistics are computed from the training split, checkpoint selection uses validation MAE, and MAE and RMSE are evaluated in physical units under each dataset's established mask.

4.4 MAIN RESULTS

Table 2: Highway forecasting performance (MAE/RMSE; lower is better).

	PeMS03		PeMS04		PeMS07		PeMS08		PeMSD7(M)		PeMSD7(L)	
Model	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
ARIMA	34.92	46.73	32.91	47.42	37.56	58.17	30.67	43.51	7.27	13.20	7.51	12.39
FC-LSTM	20.85	34.18	25.93	39.27	29.12	44.82	22.31	34.16	4.16	7.51	4.66	8.20
GRU-ED	18.67	31.84	22.89	38.15	26.94	42.53	21.38	35.27	4.78	9.05	3.98	7.71
DCRNN	17.73	29.97	20.94	32.91	24.89	38.02	16.58	25.93	3.83	7.18	4.33	8.33
ASTGCN	16.98	29.01	22.41	34.52	23.51	37.15	17.92	27.78	3.61	6.87	4.09	7.64
STGCN	17.16	29.78	20.67	34.11	24.78	38.52	17.16	26.58	4.01	7.55	4.84	8.28
STGODE	16.18	27.16	20.36	32.08	22.04	36.71	16.35	25.37	2.97	5.66	3.22	5.98
STSGCN	17.09	28.56	20.69	32.91	23.72	38.14	16.76	26.22	3.01	5.93	3.61	6.88
MSTGCN	19.12	31.24	23.39	36.41	28.31	42.78	18.67	28.64	3.54	6.14	3.58	6.43
TFormer	**14.82**	26.18	19.47	32.35	21.39	35.07	15.73	25.69	2.88	5.84	3.02	5.97
PDFormer	15.13	26.78	18.73	31.24	20.47	33.94	14.03	24.29	2.81	5.60	2.92	5.90
STAEformer	14.83	25.97	**18.06**	30.97	19.87	33.18	**13.87**	24.01	2.72	5.56	2.93	5.81
STGNCDE	16.47	27.83	19.21	31.74	22.17	36.08	15.94	25.48	**2.68**	**5.39**	**2.87**	**5.76**
CaST	15.62	26.94	18.84	31.38	21.03	34.57	14.76	24.83	3.21	6.03	3.37	6.30
RAST	15.20	**25.81**	18.39	**30.55**	**19.52**	**32.97**	14.16	**23.33**	2.96	5.99	3.21	6.33
MFormer	15.88	26.61	19.12	30.98	20.45	33.41	14.70	24.49	3.22	6.13	3.44	6.51
FRAME	**14.36**	**23.45**	**18.04**	**30.26**	**18.94**	**32.40**	**13.28**	**22.95**	**2.57**	**5.25**	**2.76**	**5.66**

In Tables 2 and 3, red and blue bold mark the best and second-best values, respectively, while green highlights FRAME. FRAME obtains the lowest MAE and RMSE among the compared methods on all six $H = 12$ highway benchmarks despite node counts ranging from 170 to 1,026 and the change from traffic flow to speed. This consistency supports the first challenge: local affine fibers remove node-dependent level and scale, while whitened Galerkin coordination supplies a stable relational chart across markedly different graph sizes. The especially large PeMS03 RMSE reduction indicates fewer severe deviations, consistent with bounded robust flux and joint graph–horizon optimization rather than independently decoded future slices. On the four $H = 1$ urban-flow tasks, the temporal and curvature sums vanish by construction, leaving boundary representation, affine continuation, conservative graph exchange, and the reference–action–context readout. FRAME achieves the lowest MAE among the compared methods on every inflow and outflow task, supporting the applicability of the same extension principle to both multi-step and direct prediction. Together with the national meteorological results in Appendix C, these protocols support the complete boundary–continuation–innovation chain: measured structure is retained, innovation is coordinated in the geometry available to each task, and predictions return to node-local physical units through the same formulation.

Table 3: Urban-flow forecasting performance (MAE/RMSE; lower is better).

Model	T-Drive Inflow		T-Drive Outflow		CHIBike Inflow		CHIBike Outflow	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
DCRNN	21.67	37.84	21.29	37.42	4.15	5.86	4.12	5.71
ASTGCN	18.34	33.21	18.36	33.31	3.98	**5.09**	3.89	5.49
STGCN	20.93	37.24	20.48	36.85	4.12	5.83	4.07	5.66
STGODE	21.07	37.45	22.23	39.47	4.07	5.80	4.04	5.59
STSGCN	23.34	40.37	23.78	41.41	4.18	5.82	4.18	5.77
MSTGCN	21.70	37.95	21.63	37.58	4.12	5.83	4.13	5.74
TFormer	18.43	33.78	18.48	34.51	3.98	5.76	3.95	5.52
PDFormer	17.46	30.97	**17.15**	**30.61**	**3.81**	**5.31**	3.76	**5.29**
STAEformer	17.28	**30.84**	17.36	30.99	3.84	5.38	**3.73**	**5.27**
RAST	17.67	31.22	17.82	31.33	3.90	5.65	3.84	5.61
MFormer	**17.16**	30.92	17.20	31.09	3.87	5.47	3.76	5.35
FRAME	**16.74**	**30.65**	**16.80**	**30.81**	**3.56**	5.37	**3.52**	5.32

4.5 Ablation Study

Table 4: Ablation results on multi-step and direct forecasting tasks.

Variant	PeMS08		T-Drive Inflow	
	MAE	RMSE	MAE	RMSE
Full FRAME	**13.28**	**22.95**	**16.74**	**30.65**
w/o Affine Fiber	13.55	23.38	17.04	31.10
w/o Whitening	13.43	23.21	16.93	30.91
w/o Affine Trace	13.60	23.47	17.07	31.17
w/o Variational Solver	13.73	23.71	17.18	31.31

Table 4 verifies that the boundary, continuation, and innovation stages contribute complementary structure rather than interchangeable capacity. Every removal worsens both MAE and RMSE on PeMS08 and T-Drive Inflow, and removing the variational solver produces the largest degradation on both datasets, showing that unconstrained drive and context cannot replace joint coordination by the action. Removing the affine trace causes the next strongest loss because the innovation branch must rediscover level and local trajectory that can instead be continued directly from observation. The affine fiber and whitening also yield consistent gains: the former preserves the node-local map between standardized dynamics and physical units, whereas the latter stabilizes sample-conditioned relational projection. This ordering remains visible for one-step T-Drive, where temporal and curvature terms are absent, so the boundary and graph-side benefits are not artifacts of multi-horizon smoothing. Appendix B further isolates the graph, temporal, and curvature terms, identifies four solver steps as the strongest operating point, and uses action-consistency analysis to show that the learned iterations progressively approach the same explicit variational target.

5 Conclusion

We presented FRAME, a boundary-conditioned formulation that rethinks spatio-temporal forecasting as the reconstruction of a future field from an observed boundary over the graph–horizon product domain. It closes the boundary–continuation–innovation chain: affine fibers retain node-local physical coordinates, whitened Galerkin projection constructs a stable sample-conditioned relational chart, a constant-preserving trace carries measured level and trajectory, and a strongly convex robust action coordinates innovation through conservative graph exchange, temporal variation, curvature, and boundary attachment. The reference–action–context readout restores physical-unit predictions. Theory establishes affine factorization, non-expansive projection, a unique variational target, and bounded conservative influence. Across twelve highway, urban-flow, and meteorological tasks, FRAME remains effective across graph scales, signals, horizons, and multi-step, direct, and long-range protocols. Ablations recover the boundary, continuation, and solver hierarchy; term and depth studies show complementary structural roles; and numerical analysis connects the learned iterations to the explicit optimization target. FRAME establishes action-minimizing extension as a forecasting principle: observation defines the boundary, measured structure is continued, and a coordinated variational field determines future innovation.

Reproducibility Statement

Appendix A provides the operator formulation and theoretical proofs. Appendix B reports action-term ablations, solver-depth sensitivity, and solver-to-action consistency verification. Appendix C provides the additional meteorological forecasting results.

AI Use Statement

Generative AI tools were used to assist with reasoning, language editing, and manuscript organization. The authors reviewed and verified all AI-assisted text, mathematical statements, citations, implementation details, and experimental claims, and remain responsible for the complete content of the submission.

References

Brandon Amos and J. Zico Kolter. OptNet: Differentiable optimization as a layer in neural networks. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 136–145. PMLR, 2017. URL https://proceedings.mlr.press/v70/amos17a.html.

Lei Bai, Lina Yao, Can Li, Xianzhi Wang, and Can Wang. Adaptive graph convolutional recurrent network for traffic forecasting. In *Advances in Neural Information Processing Systems*, volume 33, pp. 17804–17815, 2020.

George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel, and Greta M. Ljung. *Time Series Analysis: Forecasting and Control*. Wiley, 5 edition, 2015.

Shuhao Cao. Choose a transformer: Fourier or Galerkin. In *Advances in Neural Information Processing Systems*, volume 34, pp. 24924–24940. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/d0921d442ee91b896ad95059d13df618-Paper.pdf.

Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using RNN encoder–decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1724–1734, 2014. doi: 10.3115/v1/D14-1179.

Jeongwhan Choi, Hwangyong Choi, Jeehyun Hwang, and Noseong Park. Graph neural controlled differential equations for traffic forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pp. 6367–6374, 2022. doi: 10.1609/aaai.v36i6.20587.

Zheng Fang, Qingqing Long, Guojie Song, and Kunqing Xie. Spatial-temporal graph ODE networks for traffic flow forecasting. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pp. 364–373, 2021. doi: 10.1145/3447548.3467430.

Yisong Fu, Fei Wang, Zezhi Shao, Boyu Diao, Lin Wu, Zhulin An, Chengqing Yu, Yujie Li, and Yongjun Xu. On the integration of spatial-temporal knowledge: A lightweight approach to atmospheric time series forecasting. In *Advances in Neural Information Processing Systems*, volume 38, pp. 82509–82538. Curran Associates, Inc., 2025. doi: 10.52202/085713-2760. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/7691484a7a35d5e2742279c1d926b778-Paper-Conference.pdf.

Shengnan Guo, Youfang Lin, Ning Feng, Chao Song, and Huaiyu Wan. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pp. 922–929, 2019. doi: 10.1609/aaai.v33i01.3301922.

Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9(8): 1735–1780, 1997. doi: 10.1162/neco.1997.9.8.1735.

Puqing Hu, Chunjiang Wu, Chen Wang, Xin Yang, Zhibin Li, Tinghui Chen, and Shijie Zhou. Memory-augmented spatio-temporal transformer for robust traffic flow forecasting. *Biomimetics*, 11(3):170, 2026. doi: 10.3390/biomimetics11030170.

Jiawei Jiang, Chengkai Han, Wayne Xin Zhao, and Jingyuan Wang. PDFormer: Propagation delay-aware dynamic long-range transformer for traffic flow prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 4365–4373, 2023. doi: 10.1609/aaai.v37i4.25556.

Nikola Kovachki, Zongyi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural operator: Learning maps between function spaces with applications to PDEs. *Journal of Machine Learning Research*, 24(89):1–97, 2023. URL http://jmlr.org/papers/v24/21-1524.html.

Yaguang Li, Rose Yu, Cyrus Shahabi, and Yan Liu. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. In *International Conference on Learning Representations*, 2018.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations*, 2021. URL https://arxiv.org/abs/2010.08895.

Hangchen Liu, Zheng Dong, Renhe Jiang, Jiewen Deng, Jinliang Deng, Quanjun Chen, and Xuan Song. Spatio-temporal adaptive embedding makes vanilla transformer SOTA for traffic forecasting. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, pp. 4125–4129, 2023. doi: 10.1145/3583780.3615160.

Vishal Monga, Yuelong Li, and Yonina C. Eldar. Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing. *IEEE Signal Processing Magazine*, 38(2):18–44, 2021. doi: 10.1109/MSP.2020.3016905.

Weilin Ruan, Xilin Dang, Ziyu Zhou, Sisuo Lyu, and Yuxuan Liang. A retrieval augmented spatio-temporal framework for traffic prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 39163–39172, 2026. doi: 10.1609/aaai.v40i46.41264.

Chao Song, Youfang Lin, Shengnan Guo, and Huaiyu Wan. Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pp. 914–921, 2020. doi: 10.1609/aaai.v34i01.5438.

Yutong Xia, Yuxuan Liang, Haomin Wen, Xu Liu, Kun Wang, Zhengyang Zhou, and Roger Zimmermann. Deciphering spatio-temporal graph forecasting: A causal lens and treatment. In *Advances in Neural Information Processing Systems*, volume 36, pp. 37068–37088, 2023. doi: 10.52202/075280-1611.

Haoyang Yan, Xiaolei Ma, and Ziyuan Pu. Learning dynamic and hierarchical traffic spatiotemporal features with transformer. *IEEE Transactions on Intelligent Transportation Systems*, 23(11): 22386–22399, 2022. doi: 10.1109/TITS.2021.3102983.

Bing Yu, Haoteng Yin, and Zhanxing Zhu. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, pp. 3634–3640, 2018. doi: 10.24963/ijcai.2018/505.

A Formal Properties of Boundary-Conditioned Future Extension

This appendix gives the operator form and proofs underlying Section 3.7. The statements concern exactly the coordinate construction, whitened projection, affine trace, and robust action used by FRAME.

A.1 Product-Lattice Operator Form

Choose an orientation for the conservative support edges $\mathcal{E}_c$ and let $B_c \in \mathbb{R}^{|\mathcal{E}_c| \times N}$ be the resulting incidence matrix. Let $B_H \in \mathbb{R}^{\max(H-1,0)\times H}$ be the forward first-difference matrix, $C_H \in \mathbb{R}^{\max(H-2,0)\times H}$ the second-difference matrix, and $e_1 \in \mathbb{R}^H$ the first coordinate vector. Stacking A in (h, i, k) order gives $\mathbf{a} = \mathrm{vec}(A) \in \mathbb{R}^{HNK}$. The four structural operators are

$$\begin{aligned} \mathsf{D}_G &= I_H \otimes B_c \otimes I_K, \quad & \mathsf{D}_t &= B_H \otimes I_N \otimes I_K, \\ \mathsf{D}_c &= C_H \otimes I_N \otimes I_K, \quad & \mathsf{E}_1 &= e_1^\top \otimes I_N \otimes I_K. \end{aligned} \tag{17}$$

After absorbing the positive edge weights, gates, and channel-wise conductivities into separable convex functions $\Psi_G, \Psi_t, \Psi_c, \Psi_b$, Equation 11 becomes

$$\mathcal{J}(\mathbf{a}) = \frac{1}{2}\|\mathbf{a} - \mathbf{d}\|_2^2 + \Psi_G(\mathsf{D}_G\mathbf{a}) + \Psi_t(\mathsf{D}_t\mathbf{a}) + \Psi_c(\mathsf{D}_c\mathbf{a}) + \Psi_b(\mathsf{E}_1\mathbf{a} - \mathbf{b}). \tag{18}$$

Thus graph and temporal modeling are derivatives of the same unknown future field along the two axes of $\mathcal{G}_c \square P_H$. For $H < 2$, B_H is empty and the first-order temporal action vanishes; for $H < 3$, C_H is empty and the curvature action vanishes. The boundary operator selects the first future slice, making the problem an attached extension rather than a free future field.

A.2 Local Affine Factorization

Proof of Proposition 1. For $a_i > 0$, positive affine maps preserve order, so $\mathrm{median}_\tau(a_i x_{\tau i} + b_i) = a_i m_i + b_i$. The unfloored scale therefore satisfies

$$\bar{s}_i' = \frac{1}{T}\sum_\tau |a_i x_{\tau i} + b_i - (a_i m_i + b_i)| = a_i \bar{s}_i.$$

When both scale floors are inactive, $s_i = \bar{s}_i$ and $s_i' = a_i s_i$. Hence $p'_{\tau i} = (a_i x_{\tau i} + b_i - a_i m_i - b_i)/(a_i s_i) = p_{\tau i}$. First and second differences of p_i are consequently unchanged. The scale floor keeps the implemented map well defined for constant or nearly constant histories; the exact factorization applies on the nondegenerate regime stated in the proposition. □

A.3 Affine Trace Reproduction

Proposition 4 (Affine trace reproduction)**.** *For every node and horizon, the trace in Equation 7 preserves constants and is translation consistent:*

$$p_i = c\mathbf{1} \implies \rho_{hi} = c, \qquad \rho_{hi}(p_i + c\mathbf{1}) = \rho_{hi}(p_i) + c.$$

Proof. Both statements follow immediately from $\sum_\tau w_{hi\tau} = 1$. The affine constraint preserves level while allowing signed weights; it does not impose the moment conditions required for exact linear or quadratic reproduction. □

A.4 Whitened Galerkin Geometry

Proof of Proposition 2. Reduced QR factorization produces $Q \in \mathbb{R}^{N\times r}$ with $Q^\top Q = I_r$. Since $\Phi = \sqrt{N}Q$, $N^{-1}\Phi^\top\Phi = I_r$ and $\mathcal{P}_\Phi = N^{-1}\Phi\Phi^\top = QQ^\top$. Consequently, $\mathcal{P}_\Phi^\top = \mathcal{P}_\Phi$ and $\mathcal{P}_\Phi^2 = Q(Q^\top Q)Q^\top = \mathcal{P}_\Phi$. Its eigenvalues lie in $\{0, 1\}$, so its spectral norm is one and $\|\mathcal{P}_\Phi V\|_F \le \|V\|_F$. □

This property applies to the Galerkin analysis–synthesis map itself. Learned residual gates and point-wise transformations act after that controlled projection.

A.5 Temporal Geometry

Proposition 5 (Difference null spaces)**.** *For a horizon sequence $q \in \mathbb{R}^H$, $B_H q = 0$ if and only if q is constant when $H \geq 2$. Moreover, $C_H q = 0$ if and only if $q_h = c_0 + c_1 h$ for constants c_0, c_1 when $H \geq 3$.*

Proof. For $H \geq 2$, the first statement follows because every adjacent difference must vanish. For $H \geq 3$, $C_H q = 0$ makes consecutive first differences equal; their common value is c_1, and summation gives $q_h = c_0 + c_1 h$. The converse directions follow by direct substitution. □

The first-order action penalizes state change itself. The curvature action penalizes change of state change and vanishes on persistent horizon-wise velocity, so a nonzero linear trend incurs no additional curvature cost even though its slope remains visible to the first-order term.

A.6 Unique, Bounded, and Conservative Innovation

Proof of Theorem 1. For one channel of the robust potential,

$$\frac{d^2}{dv^2}\left[\frac{\kappa}{\beta}\log\cosh(\beta v)\right] = \kappa\beta\,\mathrm{sech}^2(\beta v) \geq 0.$$

Hence every robust potential is convex. Positive weighted sums preserve convexity, and composition with the linear operators in Equation 17 preserves convexity. The fidelity term $\frac{1}{2}\|\mathbf{a}-\mathbf{d}\|_2^2$ is 1-strongly convex, so their sum is 1-strongly convex. The same quadratic term makes $\mathcal{J}$ coercive. A continuous coercive function on finite-dimensional Euclidean space attains a minimizer, and strong convexity makes that minimizer unique. □

Proof of Proposition 3. Component-wise differentiation gives

$$\frac{d}{dv_k}\frac{\kappa_k}{\beta_k}\log\cosh(\beta_k v_k) = \kappa_k \tanh(\beta_k v_k),$$

whose magnitude is at most κ_k. For graph edge $\{i, j\} \in \mathcal{E}_c$, the effective bound additionally contains its fixed weight and positive boundary-conditioned gate. The implementation scatters the same edge flux with opposite signs to its two endpoints. Equivalently, graph coordination has the incidence form $B_c^\top F^G$. Because $B_c\mathbf{1} = 0$, $\mathbf{1}^\top B_c^\top F^G = 0$, proving zero node-sum graph contribution at every horizon and channel. □

Strong convexity defines $A^\star$ independently of the number of unrolled steps. Equation 13 uses the learned drive as initialization and applies S shared preconditioned gradient steps, producing the finite-depth approximation consumed by the readout. The affine trace carries inherited structure, the reduced action correction decodes structurally coordinated innovation, and the full-width context correction supplies complementary predictive detail.

B Variational Solver Ablation and Sensitivity

B.1 Action-Term Ablation

Table 5: Ablation of the structural terms in the variational solver.

Variant	PeMS08		PeMSD7(L)	
	MAE	RMSE	MAE	RMSE
w/o Graph Term	13.52	23.34	2.84	5.79
w/o Temporal Term	13.46	23.27	2.82	5.75
w/o Curvature Term	13.39	23.16	2.79	5.71
Full FRAME	**13.28**	**22.95**	**2.76**	**5.66**

Table 5 isolates the three derivative terms that coordinate the future action field. Removing graph exchange causes the largest degradation on both datasets, showing that conservative relational redistribution provides the strongest structural correction. Removing first-order temporal variation also consistently increases MAE and RMSE because adjacent future slices are no longer directly coordinated. The curvature term contributes a smaller but systematic gain by regulating changes in horizon-wise velocity. Full FRAME performs best in every column, confirming that graph exchange, temporal variation, and curvature provide complementary constraints rather than interchangeable regularizers.

B.2 Sensitivity to Solver Steps

Table 6: Sensitivity to the number of variational solver steps on PeMS08.

Solver Steps S	MAE	RMSE
1	13.47	23.24
2	13.35	23.07
4	**13.28**	**22.95**
6	13.29	22.97
8	13.31	23.00

Table 6 shows a clear improvement as the unrolled solver deepens from one to four steps, indicating that repeated preconditioned updates progressively refine the coordinated action field. Performance reaches its optimum at $S = 4$; six and eight steps remain close but do not improve either metric. We therefore use $S = 4$, which provides the strongest predictive finite-depth realization in this sensitivity study.

B.3 Solver-to-Action Consistency Verification

We freeze the trained FRAME checkpoints and record $A^{(0)}, \ldots, A^{(4)}$ together with every condition of Equation 11 on 256 uniformly spaced validation windows from PeMS08 and PeMSD7(L). For each window, a float64 L-BFGS solve with strong-Wolfe line search provides a high-accuracy reference $A^{\star}_{\mathrm{num}}$. We report the normalized action gap $G_J^{(s)}$, distance $E_A^{(s)}$ to this numerical minimizer, and relative stationarity residual $R_{\nabla}^{(s)}$, each normalized to one at $A^{(0)}$.

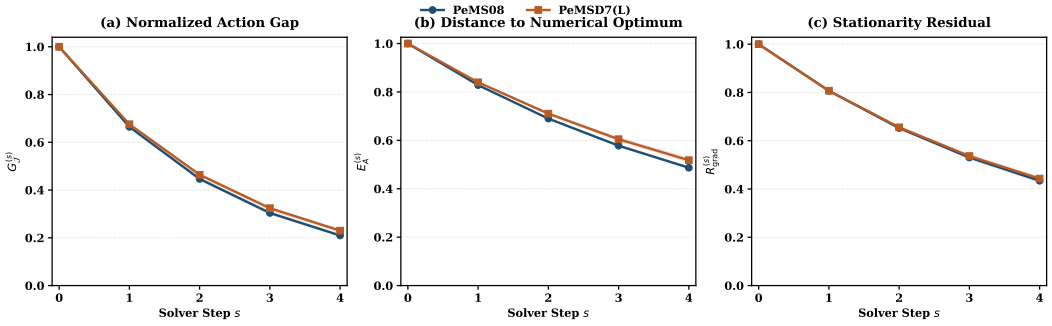

Figure 2: FRAME solver states progressively approach the numerical action minimizer.

Table 7: Finite-step consistency with the numerical variational target.

Dataset	Descent rate	$G_J^{(4)}$	$E_A^{(4)}$	$R_{\nabla}^{(4)}$	Prediction difference
PeMS08	100%	0.210	0.487	0.434	0.104%
PeMSD7(L)	100%	0.230	0.518	0.443	0.041%

Figure 2 directly verifies that the learned solver trajectory follows the stated variational action toward its numerical minimizer. The implementation gradient matches automatic differentiation to relative errors below 2.5×10^{-16}, and every transition from step zero through step four decreases the stated action on all evaluated windows. Four learned steps remove 79.0% and 77.0% of the initial action suboptimality on PeMS08 and PeMSD7(L), respectively, while the distance to $A^{\star}_{\mathrm{num}}$ and the first-order stationarity residual decrease at every step. The numerical targets themselves reach mean stationarity residuals of 1.8×10^{-9} and 2.1×10^{-9}. Replacing $A^{(4)}$ by $A^{\star}_{\mathrm{num}}$ changes the final predictions by only 0.104% and 0.041%, demonstrating that four learned steps already reproduce the prediction induced by the converged variational target with negligible discrepancy.

C Additional Meteorological Forecasting Results

ChinaTemp and ChinaWind contain daily averaged temperature and wind-speed observations from 382 and 396 stations across China, respectively, covering 2013–2022. Both tasks use the previous 60 days to forecast the next 30 days (Fu et al., 2025). Tables 8 and 9 report MAE and RMSE in the corresponding physical variables; red and blue bold identify the best and second-best values, and green highlights FRAME.

Table 8: ChinaTemp forecasting performance (lower is better).

Model	MAE ↓	RMSE ↓
ARIMA	4.026	5.396
AGCRN	2.999	4.336
MTGNN	3.168	4.221
GTS	3.082	4.298
Informer	3.045	4.477
FEDformer	3.455	4.665
DSformer	3.605	4.919
PatchTST	3.600	4.909
iTransformer	3.193	4.292
DLinear	3.659	4.990
Corrformer	3.377	4.728
STELLA-10k	3.003	4.167
STELLA-opt	**2.975**	**4.112**
FRAME	**2.882**	**4.028**

Table 9: ChinaWind forecasting performance (lower is better).

Model	MAE ↓	RMSE ↓
AGCRN	5.061	7.458
MTGNN	5.055	7.294
GTS	4.997	7.312
Informer	5.279	7.832
FEDformer	5.051	7.334
PatchTST	5.033	7.295
iTransformer	4.995	7.248
Corrformer	4.950	7.224
STELLA-10k	4.876	7.118
STELLA-opt	**4.869**	**7.104**
FRAME	**4.793**	**7.037**

The meteorological results extend FRAME from transportation networks to nationwide station fields and from short-horizon prediction to a 60-to-30 long-range regime. On ChinaTemp, FRAME improves the strongest comparison from 2.975/4.112 to 2.882/4.028 MAE/RMSE; on ChinaWind, it improves 4.869/7.104 to 4.793/7.037. The consistent reductions in both absolute and squared error support the same method-level account observed on transportation data: local affine fibers reconcile station-dependent physical scales, whitened coordination forms a stable sample-conditioned relational chart, and the variational action organizes future innovation over the longer horizon rather than decoding each future slice independently. The gains across temperature and wind speed therefore connect the boundary–continuation–innovation formulation to distinct physical variables and nationwide relational geometries.