# EMBER: EVIDENCE-DRIVEN MODAL BOUNDARY EVOLUTION FOR SPATIO-TEMPORAL FORECASTING

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Spatio-temporal forecasting requires a coordinated transformation from an observed relational field to a future whose spatial organization, regime evolution, temporal memory, and physical scale remain mutually consistent. We introduce *EMBER* (Evidence-driven Modal Boundary Evolution and Readout), which formulates this task as a coupled boundary-to-future operator. The history and graph first form a coordinated boundary state; this state conditions a node-wise rupture geometry, direct multi-horizon evolution, and a sample-level fractional memory law, while a robust boundary fiber restores local physical units. Our dependency-aligned theoretical analysis establishes stable Galerkin coordination, exact rupture decomposition with a retained orthogonal complement, a quantifiable evidence barrier governing modal reversal and coordinate-wise dynamic regimes, and positive power-law memory with sublinear accumulation and scale-covariant reconstruction. These results show that EMBER preserves boundary information while evolving it through controlled history-dependent dynamics and returning forecasts consistently to local physical units. Mechanism diagnostics further confirm well-conditioned and active Galerkin coordinates together with boundary-dependent fractional orders. Across thirteen graph, grid, and weather tasks, including temperature, humidity, and 12-hour wind-speed forecasting, EMBER achieves the strongest overall performance among the compared methods.

## 1 INTRODUCTION

Many forecasting problems observe a field evolving over related entities: power-network components, environmental stations, industrial units, communication links, or spatial cells. Relations govern how disturbances propagate, yet the future need not be a smooth continuation of the recent past. A system may preserve a regime, dissipate it, or depart from it after sufficient directional evidence accumulates. We call the latter event a *rupture*: a persistent departure of the latent boundary state along a small set of sample-conditioned directions. This motivates a general operator-learning view: infer the complete future field from a finite boundary history and a relational structure.

Diffusion recurrent models couple graph transport with temporal state transitions (Li et al., 2018). Adaptive graph learners infer latent relations rather than relying only on a prescribed topology (Bai et al., 2020). Transformer forecasters use data-adaptive spatio-temporal embeddings for global interaction (Liu et al., 2023). Their success notwithstanding, repeated propagation and generic decoding do not directly encode how a measured boundary condition reorganizes into a future regime. **Challenge I: relational coordination is not synonymous with smoothing.** Directional, signed, and long-range modes can be attenuated by repeated stochastic mixing; a global representation should preserve them while adapting to the current sample. **Challenge II: boundary states can persist, decay, or rupture.** Autoregressive decoders accumulate error, whereas direct multi-horizon evolution should accumulate evidence, inject forcing, and express amplification or contraction while generating each horizon from the observed boundary. **Challenge III: memory and measurement scale are state dependent.** The same normalized response can imply different physical changes across nodes, and the conversion from latent response to displacement may exhibit fractional memory with non-integer temporal accumulation (Podlubny, 1998).

We address these challenges with *EMBER*: Evidence-driven Modal Boundary Evolution and Readout. Its central object is a coupled boundary-to-future map in which the observed history and relational



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.

structure jointly define the coordinated state from which future geometry, evolution, and memory are generated. The observed history, local boundary fiber, calendar phase, node identity, and directed graph geometry first form a globally coordinated boundary state. That same state then conditions the node-wise rupture frame, initializes an exact modal–residual decomposition, participates in every horizon driver, and determines the sample-level fractional order used to accumulate boundary-relative responses. Relational representation therefore changes the boundary from which the future is generated, while the future operator preserves this dependence through latent evolution and physical-scale reconstruction. All horizons are produced directly, making the resulting map path dependent, non-autoregressive, and scale covariant. Our contributions are:

- We formulate graph-indexed forecasting as a coupled boundary-to-future operator with explicit structural dependencies linking relational coordinates, future evolution, memory, and physical-scale reconstruction.
- We construct a sample-conditioned whitened Galerkin chart and use its coordinated boundary to condition an exact rupture–residual decomposition. Accumulated signed evidence evolves both paths directly across all horizons, preserving the full boundary state while permitting persistence, decay, amplification, and reversal.
- We introduce a boundary-conditioned fractional readout that accumulates scalar responses to latent boundary displacement and restores node-specific scale, yielding direct path-dependent and scale-covariant forecasts.
- We establish a proof chain aligned with this computation and connect it to trained behavior through mechanism diagnostics. Across thirteen graph, grid, and weather tasks, EMBER delivers consistently leading accuracy across the evaluated domains.

## 2 Related Work

**Spatio-temporal graph forecasting.** Spatio-temporal graph forecasting couples temporal dynamics with structured dependence. DCRNN embeds directed diffusion in recurrence (Li et al., 2018). STGCN factorizes graph and temporal convolutions (Yu et al., 2018). ASTGCN adds spatial–temporal attention (Guo et al., 2019). Graph WaveNet learns adaptive connectivity (Wu et al., 2019). AGCRN makes graph and recurrent parameters node adaptive (Bai et al., 2020). STGODE models continuous-depth graph dynamics (Fang et al., 2021). PDFormer uses propagation-aware attention (Jiang et al., 2023). STAEformer emphasizes adaptive embeddings (Liu et al., 2023). EMBER instead seeds signed global coordinates from the directed graph and evolves all horizons from a coordinated boundary state.

**Neural operators and global bases.** Fourier neural operators learn global maps in a fixed spectral basis (Li et al., 2021). Neural-operator theory formalizes such models as mappings between function spaces (Kovachki et al., 2023). Galerkin attention uses learned trial coordinates (Cao, 2021). EMBER instead constructs a sample-conditioned signed basis with empirical whitening, combining relational coordinates and the current state while controlling its Gram geometry. This chart directly supports future evolution on graph-indexed fields.

**Hysteresis and fractional memory.** Hysteresis formalizes responses that depend on the accumulated path rather than only the current input (Brokate & Sprekels, 2012). Fractional calculus supplies a complementary power-law memory between local response and integer-order integration (Podlubny, 1998). Forecasting networks often approximate these effects implicitly through gates or deep temporal stacks. EMBER makes both explicit and structurally coupled: accumulated evidence controls rupture dynamics, while boundary-conditioned fractional memory aggregates responses across horizons. The same coordinated boundary therefore joins relational representation, direct future evolution, and scale-restoring readout within one operator.

## 3 Method

EMBER realizes forecasting as a coupled boundary-to-future operator. As summarized in Figure 1, the history and graph form a coordinated boundary state that jointly conditions rupture geometry, direct multi-horizon evolution, and fractional scale-covariant readout for every requested horizon.

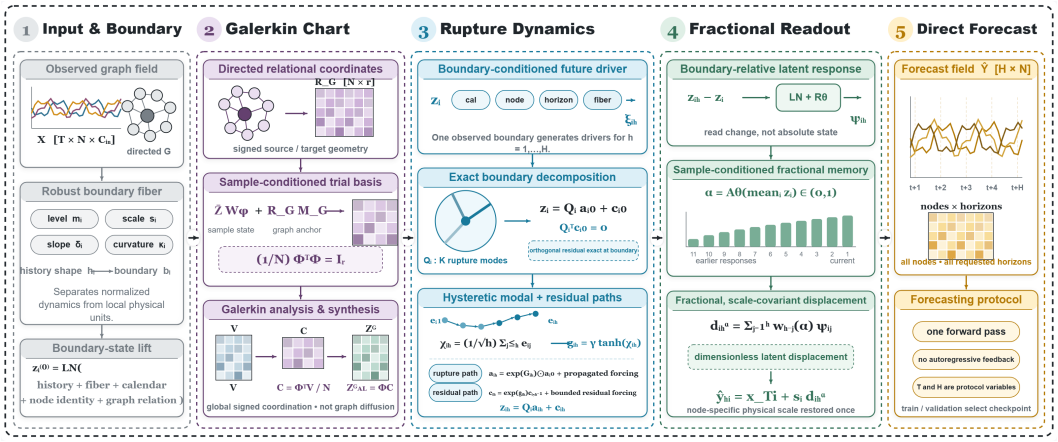


Figure 1: Overview of EMBER's boundary-to-future forecasting pipeline.

### 3.1 Problem Formulation and Notation

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a directed graph with nodes $\mathcal{V} = \{1, \dots, N\}$. For one sample, the input is $\mathbf{X} \in \mathbb{R}^{T \times N \times C_{\text{in}}}$, where $T$ is the observed length and $C_{\text{in}}$ is the number of input channels. We write $x_{\tau i} = X_{\tau,i,1}$ for the scalar forecasting variable at time $\tau$ and node $i$. The remaining channels contain known calendar indices. In the current implementation these are time of day $c_\tau^{\text{tod}}$ and day of week $c_\tau^{\text{dow}}$, shared by all nodes at observed step $\tau$. The signal channel is standardized using statistics computed from the training split only; all quantities below are in this model space unless marked "physical."

For a forecast length $H$, the target is $\mathbf{Y} = [y_{hi}] \in \mathbb{R}^{H \times N}$, with $h \in \{1, \dots, H\}$. EMBER learns

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

where $\boldsymbol{\theta}$ contains all learnable parameters. The mini-batch dimension is omitted throughout. We use $D$ for hidden width, $r$ for Galerkin rank, and $K < D$ for the number of rupture modes. The input and output lengths are protocol variables; no step assumes $T = H$. Throughout, LN denotes layer normalization, $\sigma$ is the logistic sigmoid, and $\tanh$, $\exp$, and $\mathrm{softplus}$ act component-wise on vectors. Learned neural maps act on the last tensor dimension unless stated otherwise. Importantly, $h$ indexes requested future horizons: all $H$ outputs are conditioned on the observed boundary and generated directly by $\mathcal{F}_{\boldsymbol{\theta}}$.

### 3.2 Directed Relational Coordinates

This stage constructs a fixed signed coordinate system for directed relational structure. We write

$$\mathbf{R}_{\mathcal{G}} = \mathrm{RelCoord}(\mathcal{G}) \in \mathbb{R}^{N \times r},$$

where $\mathrm{RelCoord}$ forms a directed affinity, applies a truncated singular decomposition, combines its target- and source-side directions, and normalizes columns over nodes. Its $i$-th row $\mathbf{r}_i^\top$ is the relational coordinate of node $i$. Because these coordinates are neither nonnegative nor row-stochastic, they can represent coordinated and oppositional variation. The complete construction is given in Appendix A.1.

### 3.3 Robust Boundary Fiber

**Observed local coordinates.** We use *fiber* for the node-specific coordinate system in which dynamics are interpreted relative to a robust level and scale. For node $i$, let

$$\begin{aligned} m_i &= \mathrm{median}_\tau\, x_{\tau i}, \qquad s_i = \max\left(\epsilon, T^{-1} \sum_\tau |x_{\tau i} - m_i|\right), \quad p_{\tau i} = (x_{\tau i} - m_i)/s_i, \\ \delta_{\tau i} &= \Delta p_{\tau i}, \qquad \kappa_{\tau i} = \Delta^2 p_{\tau i}, \qquad \mathbf{h}_i = [\mathbf{p}_i; \boldsymbol{\delta}_i; \boldsymbol{\kappa}_i] \in \mathbb{R}^{3T}, \\ & \mathbf{b}_i = [m_i, \log s_i, \delta_{Ti}, \kappa_{Ti}]^\top \in \mathbb{R}^4. \end{aligned} \tag{2}$$

Here $\mathbf{p}_i = [p_{1i}, \ldots, p_{Ti}]^\top$, $\boldsymbol{\delta}_i = [\delta_{1i}, \ldots, \delta_{Ti}]^\top$, and $\boldsymbol{\kappa}_i = [\kappa_{1i}, \ldots, \kappa_{Ti}]^\top$. The constant $\epsilon > 0$ prevents division by zero, and the initial differences are zero padded as detailed in Appendix A.2. Thus $\mathbf{h}_i$ records how the normalized trajectory arrived at the boundary, whereas $\mathbf{b}_i$ records the physical coordinate system and terminal derivatives.

**Boundary-state initialization.** Let $E_{\text{hist}} : \mathbb{R}^{3T} \to \mathbb{R}^D$ and $E_{\text{fib}} : \mathbb{R}^4 \to \mathbb{R}^D$ be learned lifts. Let $E_{\text{tod}}$ and $E_{\text{dow}}$ be calendar lookup tables; their sum at the last observed step is $\mathbf{e}_T^{\text{cal}} \in \mathbb{R}^D$. For node $i$, let $\mathbf{e}_i^{\text{node}}, \mathbf{e}_i^{\text{chart}} \in \mathbb{R}^D$ be learned embeddings, and let $E_{\mathcal{G}} : \mathbb{R}^r \to \mathbb{R}^D$ lift its graph coordinate. The initial state is

$$\mathbf{z}_i^{(0)} = \mathrm{LN}_{\text{in}}\big(E_{\text{hist}}(\mathbf{h}_i) + E_{\text{fib}}(\mathbf{b}_i) + \mathbf{e}_T^{\text{cal}} + \mathbf{e}_i^{\text{node}} + E_{\mathcal{G}}(\mathbf{r}_i) + \mathbf{e}_i^{\text{chart}}\big) \in \mathbb{R}^D. \tag{3}$$

Stacking over nodes gives $\mathbf{Z}^{(0)} = [\mathbf{z}_1^{(0)}, \ldots, \mathbf{z}_N^{(0)}]^\top \in \mathbb{R}^{N \times D}$. In Equation 3, $E_{\text{hist}}$ encodes normalized trajectory shape, $E_{\text{fib}}$ encodes the local physical boundary, and $\mathbf{e}_T^{\text{cal}}$ identifies the observed temporal phase. The node embedding $\mathbf{e}_i^{\text{node}}$ stores persistent identity and participates in both boundary encoding and future conditioning; $E_{\mathcal{G}}(\mathbf{r}_i)$ supplies directed relational geometry; and $\mathbf{e}_i^{\text{chart}}$ is restricted to boundary-chart initialization.

### 3.4 Sample-Conditioned Whitened Galerkin Operator

The Galerkin stage transforms node-wise hidden states into a small set of global signed modes. Its trial coordinates adapt to the current sample while remaining empirically normalized, thereby providing global coordination without interpreting the graph as a transition kernel. Galerkin and rupture modes act on different axes: the former span global variation across the $N$ nodes, whereas the latter span a sample- and node-conditioned subspace of the $D$-dimensional latent state.

Let $L$ be the number of Galerkin blocks. For block $\ell \in \{0, \ldots, L-1\}$, the input is $\mathbf{Z}^{(\ell)} \in \mathbb{R}^{N \times D}$. After layer normalization, $\overline{\mathbf{Z}}^{(\ell)} = \mathrm{LN}_1^{(\ell)}(\mathbf{Z}^{(\ell)})$, the current sample and the fixed relational coordinates generate an unwhitened trial matrix

$$\widetilde{\boldsymbol{\Phi}}^{(\ell)} = \overline{\mathbf{Z}}^{(\ell)} \mathbf{W}_\phi^{(\ell)} + \gamma_{\mathcal{G}}^{(\ell)} \mathbf{R}_{\mathcal{G}} \mathbf{M}_{\mathcal{G}}^{(\ell)} \in \mathbb{R}^{N \times r}, \tag{4}$$

where $\mathbf{W}_\phi^{(\ell)} \in \mathbb{R}^{D \times r}$, $\mathbf{M}_{\mathcal{G}}^{(\ell)} \in \mathbb{R}^{r \times r}$, and $\gamma_{\mathcal{G}}^{(\ell)} > 0$ are learned. The first term makes the basis sample dependent; the second anchors it to fixed directed geometry. We choose $r \le N$. After centering, the implementation applies stabilized root-mean-square normalization with constant $\epsilon_{\text{QR}} = 10^{-5}$ and computes a reduced QR factorization in float32, producing $\widehat{\boldsymbol{\Phi}}^{(\ell)}$. Under the standard non-degeneracy condition that the normalized trial matrix has full numerical column rank, EMBER sets

$$\boldsymbol{\Phi}^{(\ell)} = \sqrt{N}\, \widehat{\boldsymbol{\Phi}}^{(\ell)}, \qquad \frac{1}{N}(\boldsymbol{\Phi}^{(\ell)})^\top \boldsymbol{\Phi}^{(\ell)} = \mathbf{I}_r. \tag{5}$$

The $\sqrt{N}$ scaling makes the empirical inner product $N^{-1}(\boldsymbol{\Phi}^{(\ell)})^\top \boldsymbol{\Phi}^{(\ell)}$ exactly identity. Consequently, the global coefficients have a graph-size-independent normalization even when datasets contain very different numbers of nodes.

The value map $\mathbf{W}_v^{(\ell)} \in \mathbb{R}^{D \times D}$ produces $\mathbf{V}^{(\ell)} = \overline{\mathbf{Z}}^{(\ell)} \mathbf{W}_v^{(\ell)} \in \mathbb{R}^{N \times D}$. Galerkin analysis and synthesis are

$$\mathbf{C}^{(\ell)} = \frac{1}{N}(\boldsymbol{\Phi}^{(\ell)})^\top \mathbf{V}^{(\ell)} \in \mathbb{R}^{r \times D}, \qquad \mathbf{Z}_{\text{gal}}^{(\ell)} = \boldsymbol{\Phi}^{(\ell)} \mathbf{C}^{(\ell)} \in \mathbb{R}^{N \times D}. \tag{6}$$

Here $\mathbf{C}^{(\ell)}$ contains global modal coefficients and $\mathbf{Z}_{\text{gal}}^{(\ell)}$ synthesizes their node-wise contribution. Equation 6 defines a sample-conditioned low-rank operator that coordinates global signed variation through the learned trial basis. To close the block-level flow in the main text, write

$$\mathbf{Z}^{(\ell+1)} = \mathcal{B}^{(\ell)}\Big(\mathbf{Z}^{(\ell)}, \mathbf{Z}_{\text{gal}}^{(\ell)}\Big),$$

where $\mathcal{B}^{(\ell)}$ denotes the gated residual operator and position-wise feed-forward update detailed in Appendix A.4. After $L$ blocks, $\mathbf{Z} = \mathrm{LN}_{\text{bd}}(\mathbf{Z}^{(L)}) \in \mathbb{R}^{N \times D}$ is the terminal boundary state, with $i$-th row $\mathbf{z}_i^\top$. Thus the Galerkin operator changes the coordinated boundary on which every downstream prediction depends. The Galerkin frame $\boldsymbol{\Phi}^{(\ell)}$ coordinates variation across nodes, while the rupture frame $\mathbf{Q}_i$ organizes directions within each $D$-dimensional latent state; $\mathbf{z}_i$ transfers the resulting node-axis coordination into the latent rupture geometry.

### 3.5 Future Fiber Connection and Driver

Let $\mathbf{k}_{ih} \in \mathbb{R}^D$ denote a bounded second-order transport of the observed fiber to horizon $h$. Terminal slope represents local tendency, curvature represents its recent change, and clipping controls the transport magnitude before it enters the learned driver. Together with future calendar and horizon embeddings, $\mathbf{k}_{ih}$ supplies the driver with the temporal identity and local directional geometry of the observed boundary. Its explicit construction and bound are given in Appendix A.3.

Let $\mathbf{e}_h^{\mathrm{cal}} \in \mathbb{R}^D$ be the embedding of the future calendar indices obtained by advancing the last observed time of day and day of week, and let $\mathbf{e}_h^{\mathrm{hor}} \in \mathbb{R}^D$ be a learned horizon embedding. A shared driver network $D_\theta : \mathbb{R}^D \to \mathbb{R}^D$ produces

$$\boldsymbol{\xi}_{ih} = D_\theta\big(\mathrm{LN}_{\mathrm{drv}}(\mathbf{z}_i + \mathbf{e}_h^{\mathrm{cal}} + \mathbf{e}_i^{\mathrm{node}} + \mathbf{e}_h^{\mathrm{hor}} + \mathbf{k}_{ih})\big) \in \mathbb{R}^D. \tag{7}$$

### 3.6 Boundary-Conditioned Rupture Evolution

The rupture mechanism operates directly on the terminal Galerkin-coordinated boundary, allowing relational context to shape the latent directions and initial coordinates of future evolution. Consequently, changes in relational context alter $\mathbf{z}_i$ and thereby condition the latent directions, initial modal coordinates, and residual state from which all future horizons evolve. This is the central coupling between relational representation and boundary dynamics.

**Rupture frame and exact decomposition.** A rupture mode is a sample-conditioned latent direction along which persistent evidence can amplify, suppress, or reverse continuation of the current boundary state. Let $\mathbf{S}_0 \in \mathbb{R}^{D\times K}$ be a learned frame seed and let $Q_\theta : \mathbb{R}^D \to \mathbb{R}^{D\times K}$ be a learned affine map followed by reshaping. Each boundary state defines

$$\mathbf{S}_i = \mathbf{S}_0 + \beta_Q \tanh(Q_\theta(\mathbf{z}_i)) \in \mathbb{R}^{D\times K}, \qquad \mathbf{Q}_i = \mathrm{MGS}(\mathbf{S}_i), \quad \mathbf{Q}_i^\top \mathbf{Q}_i = \mathbf{I}_K, \tag{8}$$

where $0 < \beta_Q < 1$ is learned and MGS is deterministic modified Gram-Schmidt. The seed $\mathbf{S}_0$ is orthogonally initialized, $Q_\theta(\mathbf{z}_i)$ deforms it for the current sample, and the implementation clamps each MGS normalization denominator below by $\epsilon_{\mathrm{MGS}} = 10^{-5}$. Under the non-degeneracy condition that $\mathbf{S}_i$ has full numerical column rank, MGS produces the orthonormal frame $\mathbf{Q}_i$ stated in Equation 8. Let $\mathbf{P}_i = \mathbf{Q}_i\mathbf{Q}_i^\top$ and $\mathbf{P}_i^\perp = \mathbf{I}_D - \mathbf{P}_i$. The observed boundary decomposes exactly as

$$\mathbf{a}_{i0} = \mathbf{Q}_i^\top \mathbf{z}_i \in \mathbb{R}^K, \qquad \mathbf{c}_{i0} = \mathbf{P}_i^\perp \mathbf{z}_i \in \mathbb{R}^D, \qquad \mathbf{z}_i = \mathbf{Q}_i\mathbf{a}_{i0} + \mathbf{c}_{i0}, \quad \mathbf{Q}_i^\top \mathbf{c}_{i0} = \mathbf{0}. \tag{9}$$

Retaining the orthogonal residual $\mathbf{c}_{i0}$ makes the observed-boundary decomposition lossless: $\mathbf{z}_i$ is reorganized exactly into regime-sensitive modal coordinates and complementary residual coordinates.

**Accumulated modal evidence.** Hysteresis enters through accumulation. An isolated piece of opposing evidence should not immediately reverse the latent regime, whereas repeatedly aligned evidence should progressively reinforce a direction. Let $E_{\mathrm{mod}}, F_{\mathrm{mod}} : \mathbb{R}^D \to \mathbb{R}^K$ be learned affine maps. They produce modal evidence $\mathbf{e}_{ih} = E_{\mathrm{mod}}(\boldsymbol{\xi}_{ih}) \in \mathbb{R}^K$ and forcing $\mathbf{f}_{ih} = F_{\mathrm{mod}}(\boldsymbol{\xi}_{ih}) \in \mathbb{R}^K$. Hysteresis is represented by a running rupture charge and bounded signed growth,

$$\boldsymbol{\chi}_{ih} = \frac{1}{\sqrt{h}} \sum_{j=1}^{h} \mathbf{e}_{ij} \in \mathbb{R}^K, \qquad \mathbf{g}_{ih} = \gamma_g \tanh(\boldsymbol{\chi}_{ih}) \in \mathbb{R}^K, \qquad 0 < \gamma_g < \gamma_{\max}, \quad \gamma_{\max} = 0.16. \tag{10}$$

The $1/\sqrt{h}$ factor limits horizon-dependent growth in charge magnitude without erasing the reinforcement created by repeated signed evidence; unlike an average, it retains a sublinear accumulation effect. The charge $\boldsymbol{\chi}_{ih}$ is the internal hysteretic memory state: the same current evidence $\mathbf{e}_{ih}$ can induce a different growth response depending on the charge accumulated along earlier horizons. This distinguishes the mechanism from independent horizon-wise gating.

Define cumulative modal growth $\mathbf{G}_{ih} = \sum_{j=1}^{h} \mathbf{g}_{ij} \in \mathbb{R}^K$. In the fixed node-specific frame $\mathbf{Q}_i$, which is shared across forecast horizons for that node and sample, the modal state follows the diagonal affine recurrence

$$\begin{aligned} &\mathbf{a}_{ih} = \exp(\mathbf{g}_{ih}) \odot \mathbf{a}_{i,h-1} + \mathbf{f}_{ih}, \qquad \mathbf{a}_{i0} = \mathbf{Q}_i^\top \mathbf{z}_i, \\ &\quad \Longrightarrow \quad \mathbf{a}_{ih} = \exp(\mathbf{G}_{ih}) \odot \mathbf{a}_{i0} + \sum_{j=1}^{h} \exp(\mathbf{G}_{ih} - \mathbf{G}_{ij}) \odot \mathbf{f}_{ij} \in \mathbb{R}^K. \end{aligned} \tag{11}$$

The implication in Equation 11 is obtained by unrolling the recurrence; exponentiation and $\odot$ are element-wise. The term $\exp(\mathbf{G}_{ih}) \odot \mathbf{a}_{i0}$ propagates the original boundary modes, while the sum propagates every forcing $\mathbf{f}_{ij}$ through all growth occurring after its injection. Each horizon depends only on the observed boundary and future-known driver, which permits direct multi-horizon generation without previous predicted observations.

**Affine residual evolution initialized in the orthogonal complement.** The residual path starts from the exact orthogonal boundary coordinate $\mathbf{c}_{i0}$ and is not reduced to another low-dimensional predictor. Mode-specific growth is reserved for the rupture subspace; the high-dimensional residual uses one scalar rate to avoid uncontrolled directional amplification. Let $G_c : \mathbb{R}^D \to \mathbb{R}$ be a learned bounded scalar map and $E_c : \mathbb{R}^D \to \mathbb{R}^D$ a learned affine map, with $\bar{g}_{ih} = G_c(\boldsymbol{\xi}_{ih}) \in \mathbb{R}$ and $\widetilde{\mathbf{e}}^c_{ih} = E_c(\boldsymbol{\xi}_{ih}) \in \mathbb{R}^D$. The projected evidence, accumulated bounded injection, residual recurrence, and reconstructed state are

$$\begin{aligned} \mathbf{e}^c_{ih} &= \mathbf{P}_i^{\perp} \widetilde{\mathbf{e}}^c_{ih}, & \mathbf{f}^c_{ih} &= \lambda_c \tanh\left( h^{-1} \sum_{j=1}^{h} \mathbf{e}^c_{ij} \right), \\ \mathbf{c}_{ih} &= \exp(\bar{g}_{ih}) \mathbf{c}_{i,h-1} + \mathbf{f}^c_{ih}, & \mathbf{z}_{ih} &= \mathbf{Q}_i \mathbf{a}_{ih} + \mathbf{c}_{ih} \in \mathbb{R}^D, \end{aligned} \tag{12}$$

where $\lambda_c > 0$ is learned. The projection initializes residual evidence in the orthogonal complement, $\mathbf{Q}_i^{\top} \mathbf{e}^c_{ih} = \mathbf{0}$. Equation 12 then evolves an adaptive residual path from this complementary initialization, producing an additive modal–residual parameterization at every horizon. Both paths are generated directly from the observed boundary and future-known drivers.

### 3.7 Fractional Scale-Covariant Readout

The rupture dynamics determine future latent states, but not how their departure from the observed boundary becomes observable displacement. EMBER maps the boundary-relative change $\mathbf{z}_{ih} - \mathbf{z}_i$ to a horizon-wise latent response, so the readout measures future displacement from the coordinated boundary. With a shared readout $R_\theta : \mathbb{R}^D \to \mathbb{R}$ and learned scale $\lambda_\psi > 0$, define

$$\psi_{ih} = \lambda_\psi \, R_\theta(\mathrm{LN}_{\mathrm{out}}(\mathbf{z}_{ih} - \mathbf{z}_i)) \in \mathbb{R}. \tag{13}$$

EMBER parameterizes one sample-conditioned order from the mean boundary $\bar{\mathbf{z}} = N^{-1} \sum_i \mathbf{z}_i$, shared by all nodes and horizons. Thus the same coordinated boundary that conditions the node-wise rupture frames also sets the system-level memory regime through its graph-level mean. The scalar response $\psi_{ih}$ retains the node-specific effect of the evolved boundary and is accumulated by the fractional operator, while node-specific physical amplitude is restored separately through $s_i$. A learned map $A_\theta$ and generalized-binomial recursion, for $\ell \in \{1, \dots, H-1\}$, define

$$\alpha = A_\theta(\bar{\mathbf{z}}) \in (\alpha_{\min}, \alpha_{\max}) \subset (0,1), \qquad w_0(\alpha) = 1, \quad w_\ell(\alpha) = w_{\ell-1}(\alpha) \frac{\alpha + \ell - 1}{\ell}. \tag{14}$$

Restricting $\alpha$ to $(0,1)$ interpolates between near-local response and first-order accumulation. The weights are causal: the current response acts directly, while earlier horizon-wise responses contribute through lag-dependent memory. Thus the fractional operator performs a causal memory-weighted aggregation of boundary-relative scalar responses. Appendix A.4 gives the bounded parameterization used in the implementation. The causal fractional displacement and the model-space forecast are

$$d^\alpha_{ih} = \sum_{j=1}^{h} w_{h-j}(\alpha) \psi_{ij} \in \mathbb{R}, \qquad \widehat{y}_{hi} = x_{Ti} + s_i d^\alpha_{ih} \in \mathbb{R}. \tag{15}$$

Small $\alpha$ emphasizes the current response; larger $\alpha$ gives more weight to accumulated earlier responses. Because the latent displacement is dimensionless and is multiplied by the node-specific robust scale $s_i$ only at readout, two nodes with the same normalized dynamics receive proportionally scaled forecasts. Conditional on the same normalized displacement, the readout is therefore covariant to positive affine changes of the node signal. The training-only global inverse transform maps $\widehat{y}_{hi}$ to physical units.

### 3.8 Learning Objective

Let $\widehat{y}_{hi}^{\text{phys}}$ and $y_{hi}^{\text{phys}}$ denote the inverse-transformed prediction and target. For a binary official validity mask $M_{hi} \in \{0, 1\}$, EMBER minimizes physical-scale masked MAE,

$$\mathcal{L}(\boldsymbol{\theta}) = \frac{\sum_{h=1}^{H} \sum_{i=1}^{N} M_{hi} \left| \widehat{y}_{hi}^{\text{phys}} - y_{hi}^{\text{phys}} \right|}{\sum_{h=1}^{H} \sum_{i=1}^{N} M_{hi}}. \tag{16}$$

## 4 Theoretical Properties

The theory follows EMBER's forward pass and establishes three properties needed by the complete boundary-to-future operator. Full statements and proofs are provided in Appendix B.

**Stable representation without boundary loss.** Whitening makes Galerkin analysis–synthesis an orthogonal projection, so it cannot amplify latent energy and it exactly preserves features represented by the learned basis (Proposition 1). Rupture provides a lossless reorganization of the coordinated boundary: every state is reconstructed exactly from its regime-sensitive modal coordinates and the retained orthogonal residual (Proposition 2).

**Path dependence with controlled reversal.** Modal growth depends on accumulated evidence, so two histories that end with the same current input can still generate different futures. Theorem 1 quantifies this memory: bounded opposing evidence must persist long enough to overcome the accumulated charge before the growth direction can reverse. Lemma 1 further shows that every horizon is obtained directly from the observed boundary together with horizon-wise growth and forcing.

**Power-law memory with scale-consistent output.** For $0 < \alpha < 1$, fractional weights are positive, decrease with lag, and retain a power-law tail. Consequently, bounded responses accumulate as $O(h^{\alpha})$ rather than linearly with horizon (Theorem 2). Restoring each node's robust location and scale then gives a positive-affine-covariant readout (Corollary 2). Thus EMBER connects stable global coordination, lossless rupture, controlled regime change, long-memory aggregation, and physical-scale reconstruction in one operator.

## 5 Experiments

### 5.1 Datasets

We use PeMS03, PeMS04, PeMS07, and PeMS08 (Song et al., 2020), PeMSD7(M) and PeMSD7(L) (Yu et al., 2018), and T-Drive and CHIBike (Hu et al., 2026). Table 1 summarizes their domains, graph sizes, sampling intervals, time ranges, and chronological splits. The highway datasets use twelve observed steps to predict twelve future steps. T-Drive and CHIBike use six observed steps to predict the next step, following their official direct-forecasting setting.

Table 1: Dataset statistics.

| Dataset | Domain | #Nodes | #Edges | Interval | Time Range | Split |
|---|---|---|---|---|---|---|
| PeMS03 | Traffic Flow | 358 | 547 | 5 min | 09/2018–11/2018 | 6:2:2 |
| PeMS04 | Traffic Flow | 307 | 340 | 5 min | 01/2018–02/2018 | 6:2:2 |
| PeMS07 | Traffic Flow | 883 | 866 | 5 min | 05/2017–08/2017 | 6:2:2 |
| PeMS08 | Traffic Flow | 170 | 303 | 5 min | 07/2016–08/2016 | 6:2:2 |
| PeMSD7(M) | Traffic Speed | 228 | 1,132 | 5 min | 05/2012–06/2012 | 6:2:2 |
| PeMSD7(L) | Traffic Speed | 1,026 | 10,150 | 5 min | 05/2012–06/2012 | 6:2:2 |
| CHIBike | Urban Flow | 270 | 1,966 | 30 min | 07/2020–09/2020 | 7:1:2 |
| T-Drive | Urban Flow | 1,024 | 7,812 | 60 min | 02/2015–06/2015 | 7:1:2 |

### 5.2 Baselines

We compare with ARIMA (Box et al., 2015), FC-LSTM (Hochreiter & Schmidhuber, 1997), GRU-ED (Cho et al., 2014), DCRNN (Li et al., 2018), ASTGCN and MSTGCN (Guo et al., 2019),

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 |
| **EMBER** | **14.31** | **23.35** | **18.06** | **29.98** | **18.99** | **32.59** | **13.40** | **23.24** | **2.57** | **5.24** | **2.75** | **5.65** |

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

| | T-Drive Inflow | | T-Drive Outflow | | CHIBike Inflow | | CHIBike Outflow | |
|---|---|---|---|---|---|---|---|---|
| Model | 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 |
| **EMBER** | **16.65** | **30.71** | **16.72** | **30.78** | **3.54** | 5.34 | **3.50** | **5.29** |

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).

### 5.3 Experimental Settings

All experiments are run on a single NVIDIA GeForce RTX 4090 GPU with 24 GB memory, 16 CPU cores, and 120 GB system memory, using PyTorch 2.1.2 and CUDA 11.8. EMBER is trained with AdamW using linear warm-up, cosine learning-rate decay, gradient clipping, automatic mixed precision, exponential moving-average weights, and validation-based early stopping. Normalization statistics are computed from the training split, checkpoint selection uses validation MAE, and MAE/RMSE are reported in physical units under each dataset's official mask. Additional mechanism diagnostics of the learned Galerkin geometry, represented latent energy, and boundary-conditioned fractional order are reported in Appendix E.

### 5.4 Experimental Results

Tables 2 and 3 show consistent predictive gains across graph-based highway networks and grid-based urban-flow systems. In the tables, red and blue bold denote the best and second-best values among the compared methods (including ties), and green highlights EMBER. On the six highway datasets, EMBER obtains the lowest MAE and RMSE in every dataset, with the PEMS04 MAE matching the

best value. Relative to the strongest alternative in each MAE column, EMBER improves PEMS03 from 14.82 to 14.31, PEMS07 from 19.52 to 18.99, PEMS08 from 13.87 to 13.40, PeMSD7(M) from 2.68 to 2.57, and PeMSD7(L) from 2.87 to 2.75. The corresponding RMSE values of 23.35, 29.98, 32.59, 23.24, 5.24, and 5.65 are also the best across all six benchmarks; the 2.46-point reduction on PEMS03 is particularly clear. Performance remains strong from the smaller PeMSD7(M) graph to the substantially larger PeMSD7(L) graph, showing that the same formulation operates effectively across network scales. On the urban-flow tasks, EMBER achieves the lowest MAE on all four T-Drive and CHIBike inflow/outflow settings: 16.65 and 16.72 on T-Drive, and 3.54 and 3.50 on CHIBike. These results are important because the urban benchmarks differ from the highway datasets in spatial discretization, sampling frequency, forecast protocol, and flow direction. Taken together, the ten tasks show that the same boundary-to-future operator transfers from sensor graphs to spatial grids, from twelve-step forecasting to direct one-step prediction, and from aggregate traffic speed to directional urban flow while preserving consistently strong predictive accuracy. Temperature, humidity, and 12-hour wind-speed results are reported in Appendix D.

### 5.5 Ablation Study

Table 4: Component ablation on PEMS08 and T-Drive Inflow.

| Variant | PEMS08 MAE | PEMS08 RMSE | T-Drive Inflow MAE | T-Drive Inflow RMSE |
|---|---|---|---|---|
| **Full EMBER** | **13.40** | **23.24** | **16.65** | **30.71** |
| w/o Boundary Fiber | 13.68 | 23.70 | 16.92 | 31.10 |
| w/o Galerkin | 14.05 | 24.15 | 17.30 | 31.55 |
| w/o Rupture/Hysteresis | 13.82 | 23.90 | 17.05 | 31.25 |
| w/o Fractional | 13.62 | 23.58 | 16.65* | 30.71* |

*For one-step T-Drive, $H = 1$ gives $d_{i1}^{\alpha} = w_0(\alpha)\psi_{i1} = \psi_{i1}$, so the metrics coincide with Full EMBER by construction.

Table 4 shows that EMBER's four mechanisms contribute complementary functions within the boundary-to-future operator. Removing the Galerkin stage causes the largest degradation on both PEMS08 and T-Drive Inflow, confirming the importance of sample-conditioned global relational coordination. Removing rupture/hysteresis produces the second-largest loss, showing that accumulated evidence and regime-sensitive modal evolution provide additional predictive structure. The boundary-fiber ablation consistently worsens both MAE and RMSE, supporting the role of local level, scale, slope, and curvature in constructing the observed boundary. On the twelve-step PEMS08 task, removing fractional accumulation further degrades both metrics, demonstrating the contribution of horizon-dependent memory. For the one-step T-Drive setting, $H = 1$ gives $d_{i1}^{\alpha} = w_0(\alpha)\psi_{i1} = \psi_{i1}$, so the fractional operator reduces directly to the current response. The consistent ordering across the two representative settings supports EMBER's intended dependency chain: Galerkin coordinates organize relational modes, the boundary fiber anchors local geometry, rupture dynamics evolve the coordinated boundary, and fractional readout accumulates its responses across multiple horizons.

## 6 Conclusion

We presented EMBER, a coupled boundary-to-future operator for graph-indexed spatio-temporal forecasting. Its main contribution is an explicit structural dependency: the history and graph define a coordinated boundary; that boundary conditions the node-wise rupture geometry, the direct evolution of modal and residual states, and the sample-level fractional memory law; the resulting scalar responses are finally restored through local physical scale. The method couples relational representation, path-dependent future evolution, and physical-scale reconstruction through a shared coordinated boundary. Theoretical analysis follows this dependency chain, establishing stable Galerkin synthesis, exact boundary reorganization, a quantified reversal barrier, coordinate-level persistence, decay and amplification, and positive power-law memory with sublinear accumulation. Mechanism diagnostics confirm well-conditioned and active Galerkin coordinates together with boundary-dependent fractional memory, while experiments across thirteen graph, grid, and weather tasks yield consistently leading accuracy across the evaluated domains. More broadly, EMBER presents spatio-temporal forecasting as the evolution of an observed geometric boundary under relational coordination.

### Reproducibility Statement

Appendix A specifies the supporting constructions, including relational coordinates, boundary conventions, future-fiber transport, and implementation parameterization. Appendix B provides the formal theoretical results and proofs. Appendices C, D, and E provide complete test metrics, cross-domain weather results, and mechanism diagnostics, respectively.

### 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

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.

Martin Brokate and Jürgen Sprekels. *Hysteresis and Phase Transitions*, volume 121 of *Applied Mathematical Sciences*. Springer, 1 edition, 2012. ISBN 9781461240488. doi: 10.1007/978-1-4612-4048-8.

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.

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.

Haitao Lin, Zhangyang Gao, Yongjie Xu, Lirong Wu, Ling Li, and Stan Z. Li. Conditional local convolution for spatio-temporal meteorological forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pp. 7470–7478, 2022. doi: 10.1609/aaai.v36i7.20711.

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.

Igor Podlubny. *Fractional differential equations: an introduction to fractional derivatives, fractional differential equations, to methods of their solution and some of their applications*, volume 198 of *Mathematics in Science and Engineering*. Academic Press, 1998. ISBN 9780125588409.

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.

Zonghan Wu, Shirui Pan, Guodong Long, Jing Jiang, and Chengqi Zhang. Graph WaveNet for deep spatial-temporal graph modeling. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, pp. 1907–1913, 2019. doi: 10.24963/ijcai.2019/264.

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.

Ling Zhao, Yujiao Song, Chao Zhang, Yu Liu, Pu Wang, Tao Lin, Min Deng, and Haifeng Li. T-GCN: A temporal graph convolutional network for traffic prediction. *IEEE Transactions on Intelligent Transportation Systems*, 21(9):3848–3858, 2020. doi: 10.1109/TITS.2019.2935152.

Xun Zhu, Yutong Xiong, Ming Wu, Gaozhen Nie, Bin Zhang, and Ziheng Yang. Weather2K: A multivariate spatio-temporal benchmark dataset for meteorological forecasting based on real-time observation data from ground weather stations. In *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics*, volume 206 of *Proceedings of Machine Learning Research*, pp. 2704–2722. PMLR, 2023. URL https://proceedings.mlr.press/v206/zhu23a.html.

# A Supporting Constructions

## A.1 Directed relational coordinates

For an edge $j \to i$ of recorded length $\ell_{j\to i} \geq 0$, define the target-source cost

$$C_{ij}^{(1)} = \begin{cases} 0, & i = j, \\ \log(1 + \ell_{j\to i}), & (j \to i) \in \mathcal{E}, \\ +\infty, & \text{otherwise.} \end{cases} \tag{17}$$

Duplicate edges retain minimum cost. One min-plus composition adds direct and two-hop context,

$$C_{ij} = \min\left( C_{ij}^{(1)}, \min_{k\in\mathcal{V}} \{C_{ik}^{(1)} + C_{kj}^{(1)}\} \right). \tag{18}$$

Let $s_{\mathcal{G}}$ be the median positive finite cost. The directed affinity is $(A_{\mathcal{G}})_{ij} = \exp(-C_{ij}/s_{\mathcal{G}})$ for finite $C_{ij}$ and zero otherwise. With $\mathbf{A}_{\mathcal{G}} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$,

$$\mathbf{R}_0 = \tfrac{1}{2}(\mathbf{U}_r + \mathbf{V}_r)\mathbf{\Sigma}_r^{1/2}, \qquad \mathbf{R}_{\mathcal{G}} = \mathrm{ColNorm}(\mathbf{R}_0). \tag{19}$$

Here $\mathrm{ColNorm}$ centers each column over nodes and divides by its root-mean-square value. Left and right singular directions encode target- and source-side structure, respectively.

## A.2 Boundary fiber conventions

Equation 2 uses observed-window differences only:

$$\delta_{\tau i} = \begin{cases} 0, & \tau = 1, \\ p_{\tau i} - p_{\tau-1,i}, & \tau > 1, \end{cases} \qquad \kappa_{\tau i} = \begin{cases} 0, & \tau \leq 2, \\ \delta_{\tau i} - \delta_{\tau-1,i}, & \tau > 2. \end{cases} \tag{20}$$

Thus no pre-window observation is required.

## A.3 Future-fiber transport

For horizon $h$, the implementation forms

$$\begin{aligned} u_{ih} &= h\delta_{Ti} + \tfrac{1}{2}h(h-1)\kappa_{Ti}, & \bar{u}_{ih} &= 3\tanh(u_{ih}/3), \\ \mathbf{b}_{ih}^{+} &= [m_i + s_i\bar{u}_{ih}, \log s_i, \tanh(\delta_{Ti} + h\kappa_{Ti}), \kappa_{Ti}]^\top, & & \\ \mathbf{k}_{ih} &= \eta_{\mathrm{fib}} \tanh\big(E_{\mathrm{fib}}(\mathbf{b}_{ih}^{+}) - E_{\mathrm{fib}}(\mathbf{b}_i)\big), & \eta_{\mathrm{fib}} &= \tfrac{1}{2}\sigma(\rho_{\mathrm{fib}}). \end{aligned} \tag{21}$$

The bounded connection enters the future driver in Equation 7, providing a controlled boundary-relative transport before the learned evolution is evaluated. Because $|\tanh(q)| < 1$ for every finite scalar $q$, the connection satisfies the component-wise and Euclidean bounds

$$|k_{ih,d}| < \eta_{\mathrm{fib}}, \qquad \|\mathbf{k}_{ih}\|_2 < \eta_{\mathrm{fib}}\sqrt{D}.$$

## A.4 Implementation parameterization

Positive scales are represented by softplus or sigmoid transforms. In particular,

$$\gamma_{\mathcal{G}}^{(\ell)} = \mathrm{softplus}(\rho_{\mathcal{G}}^{(\ell)}), \quad \beta_Q = \sigma(\rho_Q), \quad \gamma_g = 0.16\,\sigma(\rho_g), \quad \lambda_c = \mathrm{softplus}(\rho_c).$$

The abstract Galerkin block map $\mathcal{B}^{(\ell)}$ introduced in the main text is instantiated as

$$\begin{aligned} \mathbf{U}^{(\ell)} &= \mathbf{Z}^{(\ell)} + \sigma(g_{\mathrm{op}}^{(\ell)})O^{(\ell)}(\mathbf{Z}_{\mathrm{gal}}^{(\ell)}), \\ \mathbf{Z}^{(\ell+1)} &= \mathbf{U}^{(\ell)} + \sigma(g_{\mathrm{ff}}^{(\ell)})\,\mathrm{FFN}^{(\ell)}(\mathrm{LN}_2^{(\ell)}(\mathbf{U}^{(\ell)})). \end{aligned} \tag{22}$$

Here $O^{(\ell)} : \mathbb{R}^D \to \mathbb{R}^D$ is the learned operator-output map, $\mathrm{FFN}^{(\ell)} : \mathbb{R}^D \to \mathbb{R}^D$ is the position-wise feed-forward map, and $\mathrm{LN}_2^{(\ell)} : \mathbb{R}^D \to \mathbb{R}^D$ is the second block normalization; all three act row-wise

on node states. The gates $g_{\mathrm{op}}^{(\ell)}, g_{\mathrm{ff}}^{(\ell)} \in \mathbb{R}$ are learned scalars. The modal maps introduced in the main text are parameterized as

$$\mathbf{e}_{ih} = E_{\mathrm{mod}}(\boldsymbol{\xi}_{ih}) = \mathbf{W}_e \boldsymbol{\xi}_{ih} + \mathbf{b}_e, \qquad \mathbf{f}_{ih} = F_{\mathrm{mod}}(\boldsymbol{\xi}_{ih}) = \mathrm{softplus}(\rho_f)(\mathbf{W}_f \boldsymbol{\xi}_{ih} + \mathbf{b}_f).$$

Their parameters satisfy $\mathbf{W}_e, \mathbf{W}_f \in \mathbb{R}^{K\times D}$, $\mathbf{b}_e, \mathbf{b}_f \in \mathbb{R}^K$, and $\rho_f \in \mathbb{R}$. For the residual branch, the bounded scalar map $G_c$ and affine evidence map $E_c$ are

$$\bar{g}_{ih} = G_c(\boldsymbol{\xi}_{ih}) = \tfrac{1}{2}\gamma_g \tanh(\mathbf{w}_c^\top \boldsymbol{\xi}_{ih} + b_{\bar{c}}), \qquad \widetilde{\mathbf{e}}_{ih}^{c} = E_c(\boldsymbol{\xi}_{ih}) = \mathbf{W}_c \boldsymbol{\xi}_{ih} + \mathbf{b}_c.$$

Here $\mathbf{w}_c \in \mathbb{R}^D$, $b_{\bar{c}} \in \mathbb{R}$, $\mathbf{W}_c \in \mathbb{R}^{D\times D}$, and $\mathbf{b}_c \in \mathbb{R}^D$. Finally, $\lambda_\psi = \mathrm{softplus}(\rho_\psi)$ with $\rho_\psi \in \mathbb{R}$, and the fractional order used in all reported experiments is

$$\alpha = 0.03 + 0.94\,\sigma\big(\rho_\alpha + 0.35\tanh(\mathbf{w}_\alpha^\top \bar{\mathbf{z}} + b_\alpha)\big), \tag{23}$$

where $\mathbf{w}_\alpha \in \mathbb{R}^D$ and $b_\alpha, \rho_\alpha \in \mathbb{R}$, so $(\alpha_{\min}, \alpha_{\max}) = (0.03, 0.97)$.

# B Formal Theoretical Results and Proofs

## B.1 Formal Statements

**Proposition 1** (Whitened Galerkin projection)**.** *Fix a sample and block $\ell$, and assume the normalized trial matrix has full column rank. In exact arithmetic, define*

$$\mathbf{\Pi}_\Phi^{(\ell)} = \frac{1}{N}\mathbf{\Phi}^{(\ell)}(\mathbf{\Phi}^{(\ell)})^\top.$$

*Then $\mathbf{\Pi}_\Phi^{(\ell)}$ is an orthogonal projector and*

$$(\mathbf{\Pi}_\Phi^{(\ell)})^2 = \mathbf{\Pi}_\Phi^{(\ell)}, \quad \left\|\mathbf{\Pi}_\Phi^{(\ell)}\right\|_2 = 1, \quad \left\|\mathbf{Z}_{\mathrm{gal}}^{(\ell)}\right\|_F \le \left\|\mathbf{V}^{(\ell)}\right\|_F, \quad \kappa_2\left(\frac{1}{N}(\mathbf{\Phi}^{(\ell)})^\top \mathbf{\Phi}^{(\ell)}\right) = 1.$$

*Moreover, if every feature channel of $\mathbf{V}^{(\ell)}$ lies in the learned trial subspace, equivalently if $\mathbf{V}^{(\ell)} = \mathbf{\Phi}^{(\ell)}\mathbf{B}$ for some $\mathbf{B} \in \mathbb{R}^{r\times D}$, then*

$$\mathbf{\Pi}_\Phi^{(\ell)}\mathbf{V}^{(\ell)} = \mathbf{V}^{(\ell)}.$$

*Thus the sample-conditioned Galerkin synthesis itself cannot amplify latent energy, while its empirical coefficient geometry is perfectly conditioned independently of $N$ and every representable signed mode is preserved exactly.*

**Proposition 2** (Exact rupture boundary decomposition)**.** *For every node $i$ whose frame $\mathbf{Q}_i$ satisfies $\mathbf{Q}_i^\top \mathbf{Q}_i = \mathbf{I}_K$, the coordinates in Equation 9 obey*

$$\mathbf{z}_i = \mathbf{Q}_i \mathbf{a}_{i0} + \mathbf{c}_{i0}, \qquad \mathbf{Q}_i^\top \mathbf{c}_{i0} = \mathbf{0}, \qquad \|\mathbf{z}_i\|_2^2 = \|\mathbf{a}_{i0}\|_2^2 + \|\mathbf{c}_{i0}\|_2^2.$$

*Hence the rupture representation is lossless at the observed boundary: the state is reorganized exactly into a regime-sensitive component and a retained orthogonal residual.*

**Theorem 1** (Path dependence and reversal barrier)**.** *For a fixed node, mode, and horizon, two evidence histories with the same current evidence but different preceding sums produce different modal growth. Moreover, let $S_h = \sum_{j=1}^{h} e_j > 0$ for one scalar rupture mode, and suppose subsequent opposing evidence satisfies $-u \le e_{h+t} \le 0$ with $u > 0$. If $m_{\mathrm{rev}}$ is the first step at which the corresponding growth becomes negative, then*

$$m_{\mathrm{rev}} \ge \left\lfloor \frac{S_h}{u} \right\rfloor + 1.$$

*Equality is attained by persistent evidence $e_{h+t} = -u$. Thus accumulated charge imposes a quantitative evidence barrier: bounded opposition must persist long enough to overturn the current growth direction.*

**Corollary 1** (Modal persistence, decay, amplification, and forcing)**.** *For a scalar rupture coordinate $k$, Equation 11 reads*

$$a_{ih,k} = \exp(g_{ih,k}) a_{i,h-1,k} + f_{ih,k}.$$

*If $f_{ih,k} = 0$, then $g_{ih,k} = 0$ preserves the coordinate exactly. For $a_{i,h-1,k} \neq 0$, negative growth strictly contracts its magnitude and positive growth strictly amplifies it:*

$$g_{ih,k} < 0 \implies |a_{ih,k}| < |a_{i,h-1,k}|,$$
$$g_{ih,k} > 0 \implies |a_{ih,k}| > |a_{i,h-1,k}|.$$

*When $f_{ih,k} \neq 0$, additive forcing captures both reinforcement of and opposition to the propagated state.*

**Lemma 1** (Direct modal solution)**.** *The diagonal affine recurrence in Equation 11 has the unique unrolled solution*

$$\mathbf{a}_{ih} = \exp(\mathbf{G}_{ih}) \odot \mathbf{a}_{i0} + \sum_{j=1}^{h} \exp(\mathbf{G}_{ih} - \mathbf{G}_{ij}) \odot \mathbf{f}_{ij}.$$

*Consequently, every modal horizon is obtained directly from the observed boundary together with horizon-wise growth and forcing.*

**Theorem 2** (Fractional power-law memory and sublinear accumulation)**.** *For fixed $\alpha \in (0, 1)$, the kernel in Equation 14 satisfies*

$$w_\ell(\alpha) = \frac{\Gamma(\ell + \alpha)}{\Gamma(\alpha)\Gamma(\ell + 1)}, \qquad 1 = w_0 > w_1 > \cdots > 0, \qquad w_\ell(\alpha) \sim \frac{\ell^{\alpha-1}}{\Gamma(\alpha)}.$$

*If $|\psi_{ij}| \leq M$, then*

$$|d_{ih}^{\alpha}| \leq M \frac{\Gamma(h + \alpha)}{\Gamma(\alpha + 1)\Gamma(h)} = O(h^{\alpha}).$$

*Thus EMBER retains a positive power-law memory while accumulating bounded responses sublinearly. Across the fractional family, $d_{ih}^{\alpha} \to \psi_{ih}$ as $\alpha \to 0^+$ and $d_{ih}^{\alpha} \to \sum_{j=1}^{h} \psi_{ij}$ as $\alpha \to 1^-$.*

**Corollary 2** (Conditional positive-affine covariance)**.** *Consider the model-space transformation $x'_{\tau i} = a_i x_{\tau i} + \beta_i$ with $a_i > 0$. If the robust-scale floor is inactive before and after transformation, then $\mathbf{h}'_i = \mathbf{h}_i$, $m'_i = a_i m_i + \beta_i$, and $s'_i = a_i s_i$. Conditional on the same normalized latent displacement $d_{ih}^{\alpha}$, the readout satisfies*

$$\widehat{y}'_{hi} = a_i \widehat{y}_{hi} + \beta_i.$$

Together, Propositions 1–2, Theorems 1 and 2, Corollaries 1 and 2, and Lemma 1 formalize the intended roles of whitening, rupture, hysteresis, direct evolution, fractional memory, and local scale restoration. Collectively, these results establish the structural properties of EMBER's coordinated boundary-to-future operator.

### B.2 Proofs

*Proof of Proposition 1.* Suppress the block superscript. Equation 5 gives $\mathbf{\Phi}^\top \mathbf{\Phi} = N\mathbf{I}_r$. Therefore

$$\mathbf{\Pi}_\Phi^2 = \frac{1}{N^2} \mathbf{\Phi}\mathbf{\Phi}^\top \mathbf{\Phi}\mathbf{\Phi}^\top = \frac{1}{N} \mathbf{\Phi}\mathbf{\Phi}^\top = \mathbf{\Pi}_\Phi, \qquad \mathbf{\Pi}_\Phi^\top = \mathbf{\Pi}_\Phi.$$

Thus $\mathbf{\Pi}_\Phi$ is the Euclidean orthogonal projector onto $\mathrm{col}(\mathbf{\Phi})$ and, for nonzero rank, $\|\mathbf{\Pi}_\Phi\|_2 = 1$. Equation 6 can be rewritten as $\mathbf{Z}_{\mathrm{gal}} = \mathbf{\Pi}_\Phi \mathbf{V}$, so submultiplicativity gives

$$\|\mathbf{Z}_{\mathrm{gal}}\|_F \leq \|\mathbf{\Pi}_\Phi\|_2 \|\mathbf{V}\|_F = \|\mathbf{V}\|_F .$$

Finally, $N^{-1}\mathbf{\Phi}^\top \mathbf{\Phi} = \mathbf{I}_r$ has all singular values equal to one, hence its spectral condition number is one. These identities are exact in real arithmetic under the full-column-rank condition; the float32 QR implementation satisfies them up to numerical precision, as reflected by the diagnostics in Appendix E. If $\mathrm{col}(\mathbf{V}) \subseteq \mathrm{col}(\mathbf{\Phi})$, then there exists $\mathbf{B} \in \mathbb{R}^{r \times D}$ such that $\mathbf{V} = \mathbf{\Phi}\mathbf{B}$. Therefore

$$\mathbf{\Pi}_\Phi \mathbf{V} = \frac{1}{N} \mathbf{\Phi}\mathbf{\Phi}^\top \mathbf{\Phi}\mathbf{B} = \mathbf{\Phi}\mathbf{B} = \mathbf{V},$$

which proves exact preservation for every representable feature channel. □

*Proof of Proposition 2.* Using $\mathbf{a}_{i0} = \mathbf{Q}_i^\top \mathbf{z}_i$ and $\mathbf{c}_{i0} = (\mathbf{I}_D - \mathbf{Q}_i\mathbf{Q}_i^\top)\mathbf{z}_i$,

$$\mathbf{Q}_i\mathbf{a}_{i0} + \mathbf{c}_{i0} = \mathbf{Q}_i\mathbf{Q}_i^\top \mathbf{z}_i + (\mathbf{I}_D - \mathbf{Q}_i\mathbf{Q}_i^\top)\mathbf{z}_i = \mathbf{z}_i.$$

Moreover,

$$\mathbf{Q}_i^\top \mathbf{c}_{i0} = (\mathbf{Q}_i^\top - \mathbf{Q}_i^\top \mathbf{Q}_i\mathbf{Q}_i^\top)\mathbf{z}_i = \mathbf{0}.$$

The two reconstructed components are therefore orthogonal. Since $\mathbf{Q}_i$ is an isometry on $\mathbb{R}^K$, $\|\mathbf{Q}_i\mathbf{a}_{i0}\|_2 = \|\mathbf{a}_{i0}\|_2$, and the Pythagorean identity yields the stated energy decomposition. □

*Proof of Theorem 1.* For two histories at the same horizon with equal current evidence,

$$\boldsymbol{\chi}_{ih}^{(A)} - \boldsymbol{\chi}_{ih}^{(B)} = \frac{1}{\sqrt{h}} \sum_{j=1}^{h-1} \left( \mathbf{e}_{ij}^{(A)} - \mathbf{e}_{ij}^{(B)} \right).$$

Different preceding sums therefore give different charges. Because $\gamma_g > 0$ and component-wise $\tanh$ is strictly increasing, the resulting growth vectors are different. Hence the current growth is a function of the evidence path, not only of $\mathbf{e}_{ih}$.

For the reversal result, consider one coordinate. After $m$ opposing steps,

$$S_{h+m} = S_h + \sum_{t=1}^{m} e_{h+t} \geq S_h - mu.$$

Whenever $m \leq S_h/u$, the right-hand side is nonnegative. Since division by $\sqrt{h+m}$ and $\tanh$ preserve sign, the growth coordinate cannot yet be negative. The first possible strictly negative step must therefore satisfy

$$m \geq \lfloor S_h/u \rfloor + 1.$$

If every subsequent evidence value equals $-u$, then $S_{h+m} = S_h - mu$, and this lower bound is attained. □

*Proof of Corollary 1.* Fix a scalar coordinate $k$. With $f_{ih,k} = 0$,

$$|a_{ih,k}| = \exp(g_{ih,k})|a_{i,h-1,k}|.$$

If $g_{ih,k} = 0$, the multiplicative factor is one. If $g_{ih,k} < 0$, it lies strictly between zero and one; if $g_{ih,k} > 0$, it is strictly greater than one. The corresponding magnitude inequalities are strict whenever $a_{i,h-1,k} \neq 0$. For $f_{ih,k} \neq 0$, the recurrence shows that forcing is added to the propagated coordinate, but its net effect can have either sign because cancellation is possible. □

*Proof of Lemma 1.* The claim follows by induction over $h$. For $h = 1$, $\mathbf{G}_{i1} = \mathbf{g}_{i1}$ and the closed form reduces to the recurrence. Assume the expression holds at $h-1$. Substitution into $\mathbf{a}_{ih} = \exp(\mathbf{g}_{ih}) \odot \mathbf{a}_{i,h-1} + \mathbf{f}_{ih}$ gives

$$\begin{aligned}\mathbf{a}_{ih} &= \exp(\mathbf{g}_{ih} + \mathbf{G}_{i,h-1}) \odot \mathbf{a}_{i0} \\ &\quad + \sum_{j=1}^{h-1} \exp(\mathbf{g}_{ih} + \mathbf{G}_{i,h-1} - \mathbf{G}_{ij}) \odot \mathbf{f}_{ij} + \mathbf{f}_{ih}.\end{aligned}$$

Using $\mathbf{G}_{ih} = \mathbf{G}_{i,h-1} + \mathbf{g}_{ih}$ and writing $\mathbf{f}_{ih} = \exp(\mathbf{G}_{ih} - \mathbf{G}_{ih}) \odot \mathbf{f}_{ih}$ yields the claimed sum through $j = h$. □

*Proof of Theorem 2.* Unrolling the recurrence for $w_\ell$ gives

$$w_\ell(\alpha) = \prod_{k=1}^{\ell} \frac{\alpha + k - 1}{k} = \frac{\Gamma(\ell + \alpha)}{\Gamma(\alpha)\Gamma(\ell + 1)}.$$

For $0 < \alpha < 1$, every factor is positive and

$$\frac{w_{\ell+1}(\alpha)}{w_\ell(\alpha)} = \frac{\ell + \alpha}{\ell + 1} < 1,$$

which proves strict decay. The standard Gamma-ratio asymptotic gives $\Gamma(\ell+\alpha)/\Gamma(\ell+1) \sim \ell^{\alpha-1}$, establishing power-law memory. The generalized binomial identity yields

$$\sum_{\ell=0}^{h-1} w_\ell(\alpha) = \frac{\Gamma(h+\alpha)}{\Gamma(\alpha+1)\Gamma(h)}.$$

If $|\psi_{ij}| \le M$, the triangle inequality and the change of variable $\ell = h - j$ give

$$|d_{ih}^{\alpha}| \le M \sum_{\ell=0}^{h-1} w_\ell(\alpha) = M \frac{\Gamma(h+\alpha)}{\Gamma(\alpha+1)\Gamma(h)} = O(h^{\alpha}).$$

For fixed finite $h$, $w_0 = 1$ and $w_\ell(\alpha) \to 0$ for every $\ell \ge 1$ as $\alpha \to 0^+$, whereas $w_\ell(\alpha) \to 1$ as $\alpha \to 1^-$. Applying these limits to the finite readout sum proves the two endpoint statements. □

*Proof of Corollary 2.* Let $\widetilde{s}_i = T^{-1} \sum_\tau |x_{\tau i} - m_i|$ denote the scale before flooring. Assume $\widetilde{s}_i > \epsilon$ and $a_i \widetilde{s}_i > \epsilon$, so the floor is inactive for both signals. Positive-affine equivariance of the median and absolute deviations gives

$$m_i' = a_i m_i + \beta_i, \qquad s_i' = a_i s_i.$$

Consequently,

$$p_{\tau i}' = \frac{a_i x_{\tau i} + \beta_i - a_i m_i - \beta_i}{a_i s_i} = p_{\tau i},$$

and the first and second differences are unchanged, so $\mathbf{h}_i' = \mathbf{h}_i$. Conditional on the same $d_{ih}^{\alpha}$, Equation 15 then gives

$$\widehat{y}_{hi}' = a_i x_{Ti} + \beta_i + a_i s_i d_{ih}^{\alpha} = a_i \widehat{y}_{hi} + \beta_i.$$

□

## C Complete Test Metrics

Table 5: Complete test results in physical units.

| Dataset | Protocol | MAE | RMSE | MAPE (%) |
|---|---|---|---|---|
| PEMS03 | 12→12 | 14.3055 | 23.3532 | 14.5375 |
| PEMS04 | 12→12 | 18.0582 | 29.9760 | 12.1842 |
| PEMS07 | 12→12 | 18.9926 | 32.5941 | 7.8403 |
| PEMS08 | 12→12 | 13.4029 | 23.2360 | 8.6693 |
| PEMSD7M | 12→12 | 2.5653 | 5.2386 | 6.4878 |
| PEMSD7L | 12→12 | 2.7512 | 5.6502 | 6.9976 |
| T-Drive Inflow | 6→1 | 16.6453 | 30.7113 | 13.9871 |
| T-Drive Outflow | 6→1 | 16.7185 | 30.7764 | 13.9865 |
| CHIBike Inflow | 6→1 | 3.5397 | 5.3407 | 24.8500 |
| CHIBike Outflow | 6→1 | 3.5028 | 5.2881 | 25.1527 |

## D Cross-Domain Weather Forecasting

**Dataset and protocol.** We evaluate cross-domain generalization on Weather2K-R, the public regional subset of Weather2K collected hourly from 1,866 ground stations across China. It provides 20 near-surface meteorological variables together with latitude, longitude, and altitude over 40,896 time steps (Zhu et al., 2023). Following its univariate spatio-temporal protocol, each task uses 12 hourly observations to predict the next 12 hours for air temperature, relative humidity, or wind speed. Stations are connected by a distance-based neighbor graph derived from the positional constants, and performance is measured by MAE and RMSE in the original physical units.

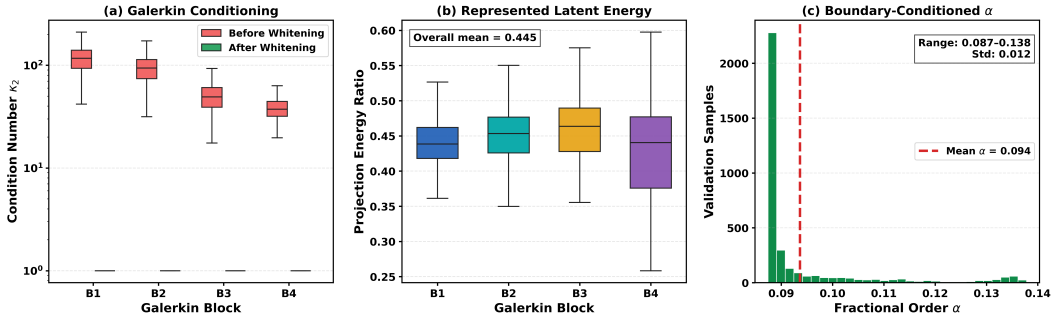


Figure 2: Mechanism diagnostics for the learned Galerkin coordinates and fractional order.

**Compared models.** The comparison covers representative convolutional and recurrent graph forecasters. DCRNN combines diffusion convolution with gated recurrent dynamics (Li et al., 2018); STGCN uses fully convolutional spatio-temporal blocks (Yu et al., 2018); and MSTGCN integrates multiple temporal components with graph convolution (Guo et al., 2019). T-GCN couples a graph convolutional network with a GRU (Zhao et al., 2020), whereas CLCRN embeds coordinate-conditioned local convolution in a recurrent encoder–decoder designed for meteorological fields (Lin et al., 2022). Table 6 summarizes their performance alongside EMBER on the three meteorological targets.

Table 6: Weather2K-R univariate forecasting performance (MAE/RMSE; lower is better).

| Model | Temperature | Humidity | Wind Speed (12h) |
|---|---|---|---|
| DCRNN | 5.0457 / 6.4070 | 16.3467 / 21.2517 | 1.39 / 1.91 |
| STGCN | 3.1297 / 4.7154 | 9.8479 / 14.5123 | 1.82 / 2.40 |
| MSTGCN | 3.6827 / 4.9195 | 14.4780 / 18.7188 | 1.50 / 2.04 |
| TGCN | 2.6429 / 3.6189 | 10.7045 / 14.3170 | 1.59 / 2.14 |
| CLCRN | **1.6498 / 2.4328** | **7.8056 / 11.1326** | **1.26 / 1.74** |
| **EMBER** | **1.5213 / 2.2136** | **6.8910 / 10.6548** | **1.13 / 1.66** |

Table 6 evaluates cross-domain forecasting over atmospheric variables with markedly different scales and dynamics. EMBER achieves the lowest MAE and RMSE in every reported setting. Relative to CLCRN, it reduces MAE by 7.8%, 11.7%, and 10.3% for temperature, humidity, and 12-hour wind speed, while reducing RMSE by 9.0%, 4.3%, and 4.6%, respectively. The consistent gains across both average and squared-error criteria support the transfer of the coupled boundary-to-future operator beyond mobility data.

## E Mechanism Diagnostics

We instrument the trained PEMS08 checkpoint on all 3,567 validation windows to examine the empirical Gram geometry, represented latent signal, and realized fractional order.

Figure 2(a) shows that whitening transforms highly variable sample-conditioned trial matrices into uniformly well-conditioned coordinates: across 14,268 block–window pairs, the mean Gram condition number decreases from 75.740 to 1.00000107. Panel (b) shows a mean represented-energy ratio of 44.5% with consistently substantial values across all four blocks, confirming that the coordinated chart actively carries latent signal. Panel (c) shows boundary-dependent fractional orders with mean 0.094 and range 0.087–0.138, emphasizing recent responses while retaining a positive power-law tail. Together, the three diagnostics verify the intended chain from stable global coordination to active latent representation and adaptive fractional readout.