CSBOO: A COHERENT SCHRÖDINGER-BRIDGE-INSPIRED ORBIT OPERATOR FOR RELIABILITY-AWARE MULTI-PERIOD SPATIO-TEMPORAL FORECASTING

Anonymous authors
Paper under double-blind review

ABSTRACT

Spatio-temporal forecasting often combines recent, daily, and weekly histories, but their reliability changes during incidents and holidays and across forecast horizons; unconstrained mixtures also make orbit weights ambiguous. We present the *Coherent Schrödinger-Bridge-Inspired Orbit Operator* (CSBOO), which replaces independent reference fusion with sequential inference over the responsibilities of three semantically anchored predictive orbits. A temporally causal backbone constructs horizon states through input-dependent directed spatial propagation, retarded attention, and known calendar phase. The retarded weights also define level, slope, and cross-period mismatch as reliability evidence. A Markov transition propagates responsibility, and a mismatch-informed Feynman–Kac potential corrects it for each sample, node, and horizon. Bounded candidate deformations and target-conditioned calibration give the normalized marginals predictive semantics; their convex barycenter yields the forecast. Across ten tasks, CSBOO achieves the best result in 18 of 20 MAE/RMSE comparisons, including 7.9% lower MAE and 4.0% lower RMSE on PeMS08 than the strongest baseline, with consistent degradation across all reported ablations.

1 INTRODUCTION

Spatio-temporal systems rarely evolve on a single clock. Traffic at 8:00 a.m. depends on the immediately preceding queue, but it may also resemble 8:00 a.m. yesterday and the same weekday one week earlier. Similar mixtures of local continuity and periodic recurrence arise in energy demand, environmental sensing, and mobility. This motivates the common practice of exposing a forecasting model to recent, daily, and weekly histories. Early multi-branch models explicitly separated closeness, period, and trend before learning a fusion (Zhang et al., 2017); later models introduced shifted periodic attention (Yao et al., 2019), adaptive graph dependencies (Wu et al., 2019), and long-range spatio-temporal attention (Jiang et al., 2023).

Periodic references are conditional on the current regime. A holiday or incident can invalidate a historical cycle locally, and the reference useful now may differ from the one useful later. Multi-period forecasting must therefore infer a coherent, horizon-dependent belief over which reference should govern each prediction. This requirement presents three challenges. **C1: Periodicity–reliability conflation.** Availability of a daily or weekly lag does not imply reliability; explicit horizon-dependent evidence must test its level and dynamics against recent context. **C2: Horizon-wise incoherence.** Independent weights switch freely, whereas global weights cannot adapt. Orbit relevance should be persistent yet switchable across horizons. **C3: Ambiguous orbit responsibility.** Because a scalar barycenter admits multiple mass–candidate decompositions, mixture weights need semantic anchors and target-conditioned calibration before they represent predictive responsibility.

CSBOO treats these horizon-wise responsibilities as a sequential latent measure. Recent, daily, and weekly references define anchored hypotheses; Markov transport models persistence and switching; and mismatch-informed Feynman–Kac correction adapts their mass to local reliability (C1–C2). Semantic anchoring and responsibility calibration give the resulting marginals predictive meaning (C3).

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.

Our contributions are:

- We formulate multi-period forecasting as sequential inference over a three-state predictive-responsibility measure rather than independent fusion.
- Retarded mismatch, Markov transport, and Feynman–Kac correction make orbit responsibility horizon-aware, persistent, and locally adaptive.
- We derive the relative-entropy path optimizer and exact marginal recursion, with guarantees for normalization, masking, and monotonicity.
- Anchored candidates and target-conditioned calibration give orbit mass predictive semantics. CSBOO leads 18 of 20 benchmark comparisons, and its ablations and diagnostics verify the three prescribed mechanisms.

Formal derivations for the path optimizer and terminal-marginal recursion, normalization and masking, mismatch-to-mass monotonicity, gauge invariance, and responsibility calibration are collected in Appendices C–G.

2 Related Work

Multi-period decomposition is especially close to our setting. ST-ResNet uses separate closeness, period, and trend branches with learned regional fusion (Zhang et al., 2017). STDN observes that periodicity is not exact and uses periodically shifted attention (Yao et al., 2019). Correlation-based selection has also been used to choose informative periodic sequences (Zhu et al., 2023). CSBOO instead embeds the three temporal references as candidate observables within a shared sequential orbit process.

Schrödinger-bridge inference provides a path-space perspective in which a stochastic law is shaped relative to a reference process through entropy regularization, and is closely connected to entropy-regularized optimal transport (Léonard, 2014; Cuturi, 2013). Modern machine learning develops related ideas for generative modeling (Wang et al., 2021; Tong et al., 2024) and time-series imputation (Chen et al., 2023). Feynman–Kac potentials provide a sequential change-of-measure mechanism for Markov path laws (Del Moral, 2004). Inspired by this path-space viewpoint, CSBOO constructs a conditional three-state orbit process in which learned horizon-wise Markov transport supplies the reference dynamics, learned emission evidence and horizon-aware mismatch jointly define the running action, entropy regularizes deviation from the reference law, and the resulting normalized terminal marginals directly determine the barycentric forecast. Since normalized weights are not uniquely attributable (Jain & Wallace, 2019), CSBOO calibrates orbit mass with semantic anchors and a target-conditioned responsibility.

3 Problem Formulation

Let $\mathcal{V}$ denote a set of N spatial nodes and $\mathbf{x}_s \in \mathbb{R}^N$ the signal at time s. Given an L-step history ending at t, the task is to predict $\mathbf{Y}_{t+1:t+H} \in \mathbb{R}^{H\times N}$. All forecasting representations are temporally causal: prediction at time t uses only observations available by t, together with known future calendar variables.

We organize the available past into three reference orbits: recent (R), daily (D), and weekly (W), with periods P_{D} and P_{W}. An *orbit* is a semantically anchored predictive hypothesis induced by one of these temporal references, rather than an unconstrained latent expert. For clarity, $r_{\ell,n}$, $d_{\ell,n}$, and $w_{\ell,n}$ denote the recent, daily-aligned, and weekly-aligned historical trajectories. For forecast horizon h, the corresponding periodic anchors are $d^{+}_{h,n} = x_{t+h-P_{\mathrm{D}},n}$ and $w^{+}_{h,n} = x_{t+h-P_{\mathrm{W}},n}$ whenever these observations are causally available. We use $\eta^{\mathrm{D}}_{h,n}, \eta^{\mathrm{W}}_{h,n} \in \{0,1\}$ to indicate their availability and collect anchors and masks as $A^{+}_{h,n} = [d^{+}_{h,n}, \eta^{\mathrm{D}}_{h,n}, w^{+}_{h,n}, \eta^{\mathrm{W}}_{h,n}] \in \mathbb{R}^4$. The model receives $X^{\mathrm{hist}} \in \mathbb{R}^{L\times N\times d_{\mathrm{hist}}}$, $A^{+} \in \mathbb{R}^{H\times N\times 4}$, and known future calendar features $\Phi^{+} \in \mathbb{R}^{H\times d_\phi}$. Let $\boldsymbol{\phi}_h$ denote the known calendar feature at horizon h, and let $\boldsymbol{\xi}_h \in \mathbb{R}^{d_\xi}$ denote a learned horizon embedding. The history and forecast axes are independent, so the formulation does not require $L = H$. Exact temporal indexing, causal availability masks, and channel construction are given in Appendix B.

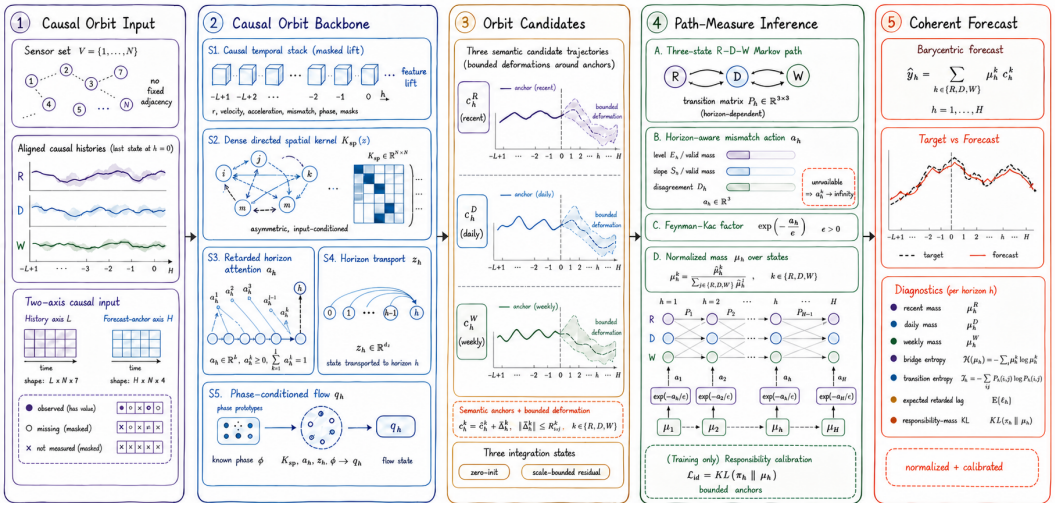

Figure 1: CSBOO overview: causal encoding, semantic orbit construction, Feynman–Kac orbit inference, and barycentric forecasting.

For orbit $k \in \mathcal{O} = \{\mathrm{R}, \mathrm{D}, \mathrm{W}\}$, the model produces a candidate $c_{h,n}^k$ and mass $\mu_{h,n}^k$. The forecast is their barycenter:

$$\widehat{y}_{h,n} = \sum_{k \in \mathcal{O}} \mu_{h,n}^k c_{h,n}^k, \qquad \boldsymbol{\mu}_{h,n} \in \Delta^{|\mathcal{O}|-1}. \tag{1}$$

Here $\boldsymbol{\mu}_{h,n} := [\mu_{h,n}^{\mathrm{R}}, \mu_{h,n}^{\mathrm{D}}, \mu_{h,n}^{\mathrm{W}}]^\top$. The learning problem is therefore to construct the orbit candidates and infer a temporally causal, normalized measure $\boldsymbol{\mu}_{1:H,n}$. We call this measure *coherent* when its successive marginals arise from one sequential path process, so changes in orbit responsibility are coupled across horizons rather than selected independently.

4 Method

Figure 1 summarizes CSBOO as sequential inference over orbit responsibilities. A shared temporally causal field constructs three anchored predictive hypotheses and evaluates their horizon-aware reliability. Markov prediction propagates prior responsibility along the forecast path, the Feynman–Kac potential corrects it with current evidence, and the resulting terminal marginals form the barycentric forecast.

4.1 Shared causal orbit backbone

Causal temporal evidence. A masked feature lift followed by left-padded dilated blocks maps the historical input to token states $U_{1:L}$ and a pre-spatial node state $s^{(0)}$:

$$(U_{1:L}, s^{(0)}) = \mathcal{E}_{\text{temp}}(X^{\text{hist}}; v), \tag{2}$$

where v is a learned node embedding. Invalid periodic values and slopes are zeroed before projection, while their masks remain explicit features. Hence the encoder is strictly causal: zero-filled invalid values cannot be mistaken for valid observations, while missingness remains explicitly represented by the masks. The feature channels, causal convolution, gating, and history readout are given in Appendix H.

Input-dependent nonreciprocal spatial kernel. Let $X \in \mathbb{R}^{N \times d}$ be the input to one spatial layer and $\bar{X} = \mathrm{LN}(X)$. For A spatial attention heads, head $m \in \{1, \dots, A\}$ has width $d_m = d/A$ and forms $Q^m = \bar{X} W_Q^m$, $K^m = \bar{X} W_K^m$, and $V^m = \bar{X} W_V^m$, where $W_Q^m, W_K^m, W_V^m \in \mathbb{R}^{d \times d_m}$. Its temperature is $\tau_m > 0$, learned independently for each head. Two separately row-normalized,

diagonal-masked kernels are

$$F_{ij}^m = \underset{j\neq i}{\mathrm{softmax}}\left(\langle Q_i^m, K_j^m\rangle/\tau_m\right),$$
$$R_{ij}^m = \underset{j\neq i}{\mathrm{softmax}}\left(\langle K_i^m, Q_j^m\rangle/\tau_m\right). \tag{3}$$

The diagonal logits are $-\infty$, so a node cannot copy itself through the message kernel. Because the two kernels are normalized by rows, R^m is not in general $(F^m)^\top$. Their odd and even combinations provide directed advection- and diffusion-like messages, producing $s = \mathcal{K}_{\mathrm{sp}}^{(L_{\mathrm{sp}})} \circ \cdots \circ \mathcal{K}_{\mathrm{sp}}^{(1)}(s^{(0)})$, where L_{sp} is the number of spatial layers. This sample-dependent nonreciprocity represents asymmetric propagation without requiring a supplied adjacency matrix; each kernel is dense over the N spatial nodes. Appendix H gives the complete message and gated residual update.

Retarded horizon field. Because the relevant historical evidence can change with forecast distance, we construct a horizon-indexed causal context rather than broadcasting one summary. We concatenate the known calendar feature and learned horizon embedding as $\psi_h = [\boldsymbol{\phi}_h; \boldsymbol{\xi}_h]$. We use two separately learned phase projections,

$$p_h = P_\phi(\psi_h), \qquad p_h^{\mathrm{qry}} = P_{\mathrm{qry}}(\psi_h),$$

where p_h conditions the decoder and emission, whereas p_h^{qry} conditions the retarded-field query. Together with s_n, the latter queries the causal token sequence:

$$g_{h,n} = \mathrm{norm}\left(W_q s_n + p_h^{\mathrm{qry}}\right),$$
$$k_{\ell,n} = \mathrm{norm}(W_k \bar{U}_{\ell,n}), \qquad h_{\ell,n}^{\mathrm{val}} = W_{\mathrm{val}} \bar{U}_{\ell,n},$$
$$\alpha_{h,n,\ell} = \mathrm{softmax}_\ell\left(\langle g_{h,n}, k_{\ell,n}\rangle/\tau_{\mathrm{delay}} + b_{h,\ell}\right), \tag{4}$$

where $\mathrm{norm}(x) = x/(\|x\|_2 + \delta)$ denotes L2 normalization, $\delta > 0$ is the shared numerical stabilizer used throughout, $\bar{U} = \mathrm{LN}(U)$, $\tau_{\mathrm{delay}} > 0$, and $b_{h,\ell}$ is a learnable relative-delay bias. The resulting delayed state is

$$\rho_{h,n} = \sum_{\ell=1}^{L} \alpha_{h,n,\ell} h_{\ell,n}^{\mathrm{val}},$$
$$u_{h,n} = \mathrm{LN}\left(s_n + \sigma(W_{\rho g}[s_n, \rho_{h,n}]) \odot \Psi_\rho([s_n, \rho_{h,n}])\right). \tag{5}$$

where $\sigma(\cdot)$ denotes the sigmoid function and Ψ_ρ is a learnable residual update map.

Horizon refinement and phase-conditioned decoder. A shared horizon-refinement spatial layer is applied independently to each delayed field:

$$z_h = \mathcal{K}_{\mathrm{sp}}^{\mathrm{hor}}(u_h), \qquad q_{h,n} = \mathcal{D}_{\mathrm{phase}}(z_{h,n}, p_h, v_n).$$

Here $\mathcal{D}_{\mathrm{phase}}$ is the shared phase-conditioned decoder detailed in Appendix H. Thus both z_h and q_h vary by sample and horizon even though their parameters are shared. The modal rotation, contraction, and residual decoder are expanded in Appendix H.

4.2 Mismatch statistics and orbit candidates

Horizon-aware, mask-normalized orbit mismatch. For $k \in \{\mathrm{D}, \mathrm{W}\}$, define the aligned periodic history by $b_{\ell,n}^{\mathrm{D}} := d_{\ell,n}$ and $b_{\ell,n}^{\mathrm{W}} := w_{\ell,n}$, and let $m_{\ell,n}^k \in \{0,1\}$ denote its validity mask. For any trajectory $x_{1:L,n}$, define its first temporal difference as $\nabla x_{1,n} = 0$ and $\nabla x_{\ell,n} = x_{\ell,n} - x_{\ell-1,n}$ for $\ell > 1$. Accordingly, ∇r and ∇b^k denote the first-order differences of the recent and aligned periodic trajectories. A slope is valid only when both endpoints are observed: $\widetilde{m}_{1,n}^k = 0$ and $\widetilde{m}_{\ell,n}^k = m_{\ell,n}^k m_{\ell-1,n}^k$ for $\ell > 1$. For feature type $\iota \in \{0,1\}$, define validity $\nu^{k,0} = m^k$, $\nu^{k,1} = \widetilde{m}^k$ and discrepancies $\zeta^{k,0} = (r - b^k)^2$, $\zeta^{k,1} = |\nabla r - \nabla b^k|$. Reusing the retarded weights from Eq. (4) aligns reliability assessment with the same horizon-specific evidence used to construct the state; both statistics are instances of

$$\mathcal{M}_{h,n}^{k,\iota} = \frac{\sum_\ell \alpha_{h,n,\ell} \nu_{\ell,n}^{k,\iota} \zeta_{\ell,n}^{k,\iota}}{\max(\sum_\ell \alpha_{h,n,\ell} \nu_{\ell,n}^{k,\iota}, \delta)}, \qquad E_{h,n}^k = \sqrt{\mathcal{M}_{h,n}^{k,0} + \delta}, \quad S_{h,n}^k = \mathcal{M}_{h,n}^{k,1}. \tag{6}$$

Here E^k detects level or regime failure, whereas S^k detects dynamical failure even when levels remain close; both assess the reliability of the corresponding horizon-aligned anchor used below. We denote the level-valid attention mass by $Z^{k,0}_{h,n} = \sum_{\ell=1}^{L} \alpha_{h,n,\ell} m^k_{\ell,n}$. The cross-period disagreement $I^{\mathrm{DW}}_{h,n}$ applies the same operator to $d - w$ under the joint daily–weekly mask and detects cross-period inconsistency. This normalization measures mismatch over valid attended history, preserving the reliability semantics under missing periodic observations. Because $\alpha_{h,n,\ell}$ depends on h, all three mismatch signals are genuinely horizon-specific. Full level, slope, disagreement, and zero-support formulas appear in Appendix I.

Local scale and anchored candidates. For the current sample ending at t, we define

$$s^{\mathrm{loc}}_{t,n} = \max\left\{ \sqrt{\frac{1}{L}\sum_{\ell=1}^{L} (\nabla r_{\ell,n})^2 + \delta},\, s_{\min} \right\}, \qquad R^{\mathrm{def}}_{t,n} = \gamma s^{\mathrm{loc}}_{t,n}, \tag{7}$$

where $\delta > 0$, $s_{\min} > 0$, and $\gamma > 0$. These are window-dependent sample–node quantities, not fixed node parameters. Let $\bar{c}^{\mathrm{R}}_{h,n} = r_{L,n}$, $\bar{c}^{\mathrm{D}}_{h,n} = d^{+}_{h,n}$, and $\bar{c}^{\mathrm{W}}_{h,n} = w^{+}_{h,n}$ denote the three semantic anchors. Candidate-specific heads H_k produce scale-bounded deformations and observables,

$$\Delta^k_{h,n} = R^{\mathrm{def}}_{t,n} \tanh H_k(s_n, z_{h,n}, q_{h,n}), \tag{8}$$

$$c^k_{h,n} = \bar{c}^k_{h,n} + \Delta^k_{h,n}. \tag{9}$$

The bounded residual keeps each candidate an explicit predictive hypothesis near its semantic temporal anchor, rather than an unconstrained expert. Candidate intervals are disjoint when $|\bar{c}^i_{h,n} - \bar{c}^j_{h,n}| > 2R^{\mathrm{def}}_{t,n}$; overlapping intervals relax this separation. Appendix I gives the exact candidate heads, while Appendix G gives the anchor-separation condition.

4.3 Feynman–Kac action and Markov transport

Let $e_{h,n} \in \mathbb{R}^{|\mathcal{O}|}$ be an emission score computed from the horizon-refined state, projected phase, and node embedding:

$$e_{h,n} = \mathrm{MLP}([z_{h,n}, p_h, v_n]), \qquad \bar{e}_{h,n} = e_{h,n} - e^{\mathrm{R}}_{h,n}\mathbf{1}. \tag{10}$$

Subtracting the recent score removes the additive softmax gauge without changing any normalized mass. The action is the gauge-fixed emission plus horizon-aware mismatch:

$$\begin{aligned} a^{\mathrm{R}}_{h,n} &= -\bar{e}^{\mathrm{R}}_{h,n} = 0, \\ a^{\mathrm{D}}_{h,n} &= -\bar{e}^{\mathrm{D}}_{h,n} + \lambda^{\mathrm{D}}_E E^{\mathrm{D}}_{h,n} + \lambda^{\mathrm{D}}_S S^{\mathrm{D}}_{h,n} + \lambda_I I^{\mathrm{DW}}_{h,n}, \\ a^{\mathrm{W}}_{h,n} &= -\bar{e}^{\mathrm{W}}_{h,n} + \lambda^{\mathrm{W}}_E E^{\mathrm{W}}_{h,n} + \lambda^{\mathrm{W}}_S S^{\mathrm{W}}_{h,n} + \lambda_I I^{\mathrm{DW}}_{h,n}, \end{aligned} \tag{11}$$

The recent orbit serves as the zero-action reference under this gauge, so the daily and weekly actions quantify their relative evidence and mismatch penalties against the recent baseline. The mismatch coefficients $\lambda^{\mathrm{D}}_E, \lambda^{\mathrm{D}}_S, \lambda^{\mathrm{W}}_E, \lambda^{\mathrm{W}}_S, \lambda_I$ are globally learned positive scalars shared across samples, nodes, and forecast horizons; each is parameterized as $\lambda = \mathrm{softplus}(\widetilde{\lambda}) > 0$. For $k \in \{\mathrm{D}, \mathrm{W}\}$, if $\eta^k_{h,n} = 0$ or $Z^{k,0}_{h,n} \le \delta$, we set $a^k_{h,n} = M \gg 1$, making its mass numerically zero. The explicit mismatch terms and the learned emission are both horizon-specific: the former through $\alpha_{h,n,\ell}$ and the latter through $(z_{h,n}, p_h)$.

Let $\epsilon > 0$ be the learned entropic temperature. For every horizon, $B_h \in \mathbb{R}^{|\mathcal{O}|\times|\mathcal{O}|}$ is a free horizon-indexed learnable transition-score matrix shared across all samples and nodes; it is not conditioned on the current sample. It is row-centered to fix its additive gauge,

$$\bar{B}_h(i,j) = B_h(i,j) - \frac{1}{|\mathcal{O}|}\sum_{k\in\mathcal{O}} B_h(i,k), \qquad \mathbf{T}_h(i,j) = \mathrm{softmax}_j\left(\bar{B}_h(i,j)/\epsilon\right). \tag{12}$$

A different kernel at each horizon acts as a persistence-and-switching prior over reference relevance (C2): its self-transitions favor continuity, while its off-diagonal transitions permit structured changes. $\mathbf{T}_h$ is global across samples and nodes; the sample-, node-, and horizon-specific corrections enter

through $a_{h,n}$. Throughout, bold $\mathbf{T}_h$ denotes the finite-state transition matrix, whereas blackboard-bold $\mathbb{P}_h$ below denotes a probability law on paths.

The initial orbit distribution is generated explicitly from the sample-specific node state:

$$o_{0,n} = H_0(s_n) \in \mathbb{R}^{|\mathcal{O}|}, \qquad \bar{o}_{0,n} = o_{0,n} - o^{\mathrm{R}}_{0,n}\mathbf{1}, \qquad \boldsymbol{\mu}_{0,n} = \mathrm{softmax}(\bar{o}_{0,n}). \tag{13}$$

Here H_0 is a shared map. Since s_n depends on the current sample and node, $\boldsymbol{\mu}_{0,n}$ is sample- and node-dependent rather than global. Each horizon then performs one Markov prediction and one Feynman–Kac correction:

$$\widetilde{\mu}^j_{h,n} = \sum_{i\in\mathcal{O}} \mu^i_{h-1,n}\mathbf{T}_h(i,j), \tag{14}$$

$$\mu^j_{h,n} = \frac{\widetilde{\mu}^j_{h,n}\exp(-a^j_{h,n}/\epsilon)}{\sum_{k\in\mathcal{O}}\widetilde{\mu}^k_{h,n}\exp(-a^k_{h,n}/\epsilon)}. \tag{15}$$

Equation (14) propagates the sequential prior over orbit responsibility; Eq. (15) is its sample-, node-, and horizon-specific Feynman–Kac evidence correction through the action potential $G^j_{h,n} = \exp(-a^j_{h,n}/\epsilon)$, which combines learned emission evidence with explicit mismatch penalties. Appendix I gives the bounded temperature parameterization and the numerically stable log-domain form used for stable evaluation. We intentionally share the same ϵ between the reference transition $\mathbf{T}_h$ and the Feynman–Kac correction, placing transport persistence and evidence sharpness on a common entropy scale.

4.4 Path free energy and barycentric forecast

Motivated by Schrödinger-bridge-inspired relative-entropy inference, CSBOO turns horizon-wise fusion into an evidence-weighted path measure that remains close to a Markov reference law. It defines a family $\{\mathbb{Q}^*_h\}_{h=1}^H$ of normalized Feynman–Kac prefix measures whose terminal marginals are recursively coupled across horizons. For notational clarity, we fix a node n and suppress the node index, writing $\mu_h(k) \equiv \mu^k_{h,n}$ and $a_h(k) \equiv a^k_{h,n}$. For prefix horizon h, define the reference path law

$$\mathbb{P}_h(k_{0:h}) = \mu_0(k_0)\prod_{s=1}^{h}\mathbf{T}_s(k_{s-1},k_s). \tag{16}$$

For $\epsilon > 0$, the prefix Gibbs measure is the unique minimizer

$$\mathbb{Q}^*_h = \arg\min_{\mathbb{Q}}\left\{\mathbb{E}_{\mathbb{Q}}\left[\sum_{s=1}^{h}a_s(k_s)\right] + \epsilon\,\mathrm{KL}(\mathbb{Q}\|\mathbb{P}_h)\right\}. \tag{17}$$

Proposition 1 (path optimizer and exact recursion). For $\epsilon > 0$, Eq. (17) has a unique Gibbs optimizer whose terminal marginals are exactly Eqs. (14)– (15). Appendix C gives the KL decomposition and forward-message proof. The resulting recursion exactly marginalizes the orbit paths, and $\boldsymbol{\mu}_{h,n}$ is the terminal marginal of the h-prefix measure.

Proposition 2 (normalization, convexity, and masking). From $\boldsymbol{\mu}_{0,n} \in \Delta^{|\mathcal{O}|-1}$, finite actions preserve normalized mass and a convex readout. Exact masking of orbit j at (h,n) is the limit $a^j_{h,n} \to +\infty$; finite M gives only numerically negligible mass. Appendix D gives the induction proof.

Proposition 3 (mismatch-to-mass monotonicity). Fix (h,n), $\epsilon > 0$, a finite action vector, and a transported prior with $\widetilde{\mu}^i_{h,n} > 0$ for every i, while holding the learned emission and every mismatch term not under variation fixed. Increasing orbit j's level or slope mismatch strictly decreases its corrected mass; increasing shared daily–weekly disagreement decreases total periodic mass. Appendix E gives the derivatives. Thus stronger periodic-failure evidence cannot increase that periodic explanation's responsibility under the stated conditions. Equation (1) returns the barycentric observable under $\boldsymbol{\mu}_{h,n}$.

4.5 ORBIT-MASS RESPONSIBILITY CALIBRATION

For the normalized path marginals to serve as predictive responsibilities, each state must retain a stable meaning. We enforce this meaning by coupling the semantic anchors and bounded deformations in Eqs. (8)–(9) with a target-conditioned responsibility during training, resolving the multiple mass–candidate decompositions admitted by Eq. (1).

Lemma 1 (additive gauge invariance). For any score vector z and scalar c, $\mathrm{softmax}(z + c\mathbf{1}) = \mathrm{softmax}(z)$; hence the two centerings select canonical logits, while barycentric ambiguity is handled separately by the semantic-anchor and responsibility-calibration mechanisms below. Appendix F gives the proof.

Using the level-valid attention mass defined above, the path action and calibration loss share the same usability mask:

$$\chi^{\mathrm{R}}_{h,n} = 1, \qquad \chi^{k}_{h,n} = \eta^{k}_{h,n}\mathbb{I}[Z^{k,0}_{h,n} > \delta], \quad k \in \{\mathrm{D}, \mathrm{W}\}. \tag{18}$$

Thus a periodic orbit enters the responsibility distribution only if its causal horizon anchor is available and the retarded history has level-valid attention mass $Z^{k,0}_{h,n} > \delta$. Let $\tau_{\mathrm{cal}} > 0$ denote the responsibility-calibration temperature. Given target $y_{h,n}$, the responsibility is

$$\pi^{k}_{h,n} = \frac{\chi^{k}_{h,n} \exp[-|y_{h,n} - c^{k}_{h,n}|/(\tau_{\mathrm{cal}} s^{\mathrm{loc}}_{t,n})]}{\sum_{j\in\mathcal{O}} \chi^{j}_{h,n} \exp[-|y_{h,n} - c^{j}_{h,n}|/(\tau_{\mathrm{cal}} s^{\mathrm{loc}}_{t,n})]}. \tag{19}$$

Collecting the orbit-wise entries gives $\boldsymbol{\pi}_{h,n} := [\pi^{\mathrm{R}}_{h,n}, \pi^{\mathrm{D}}_{h,n}, \pi^{\mathrm{W}}_{h,n}]^{\top}$. For $\tau_{\mathrm{cal}} > 0$, this is a unique, normalized statement of which anchored orbit best explains the observed target at each node and horizon. For one training sample, we optimize

$$\begin{aligned}
\mathcal{L}_{\mathrm{pred}} &= \frac{1}{HN}\sum_{h,n} \ell_{\mathrm{pred}}(\widehat{y}_{h,n}, y_{h,n}), \\
\mathcal{L}_{\mathrm{cal}} &= \frac{1}{HN}\sum_{h,n} \mathrm{KL}(\mathrm{sg}(\boldsymbol{\pi}_{h,n}) \| \boldsymbol{\mu}_{h,n}), \\
\mathcal{L}_{\mathrm{anc}} &= \frac{\sum_{h,n,k} \chi^{k}_{h,n} (\Delta^{k}_{h,n}/s^{\mathrm{loc}}_{t,n})^2}{\sum_{h,n,k} \chi^{k}_{h,n}}, \\
\mathcal{L} &= \mathcal{L}_{\mathrm{pred}} + \lambda_{\mathrm{cal}}\mathcal{L}_{\mathrm{cal}} + \lambda_{\mathrm{anc}}\mathcal{L}_{\mathrm{anc}}.
\end{aligned} \tag{20}$$

Here $\mathrm{sg}(\cdot)$ denotes the stop-gradient operator, ℓ_{pred} denotes the pointwise forecasting loss, and $\lambda_{\mathrm{cal}}, \lambda_{\mathrm{anc}} \geq 0$ control the strengths of responsibility calibration and anchor regularization, respectively. The empirical training objective averages this quantity over samples; we use MAE for ℓ_{pred} in all reported experiments, with optimization hyperparameters specified in Appendix L. The stop-gradient prevents candidate heads from chasing the responsibility target, while $\mathcal{L}_{\mathrm{anc}}$ discourages orbit-semantic drift. Targets appear only in this training loss; the forward recursion at validation and test time uses causal inputs and known phase only.

Proposition 4 (responsibility calibration consistency). For $\tau_{\mathrm{cal}} > 0$ and at least one usable orbit, Eq. (19) defines a unique $\boldsymbol{\pi}_{h,n}$ on the available support. For fixed $\boldsymbol{\pi}_{h,n}$, the calibration term $\mathrm{KL}(\boldsymbol{\pi}_{h,n} \| \boldsymbol{\mu}_{h,n})$ is uniquely minimized over the probability simplex on that support at $\boldsymbol{\mu}_{h,n} = \boldsymbol{\pi}_{h,n}$. Appendix G gives the proof and anchor qualification.

The path objective defines C1–C2 inference; the separate responsibility loss calibrates C3 during training.

5 EXPERIMENTS

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). Appendix A summarizes their variables, spatial scales, sampling intervals, time ranges, and splits.

Table 1: Forecasting accuracy on highway benchmarks (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	**29.93**	**19.52**	**32.73**	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
CSBOO (ours)	**14.75**	**25.26**	**17.98**	**29.88**	**18.99**	**32.23**	**12.78**	**22.39**	**2.58**	**5.24**	**2.86**	**5.77**

Table 2: Forecasting accuracy on urban-flow benchmarks (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	**30.54**	17.36	30.84	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
CSBOO (ours)	**17.11**	**30.21**	**17.11**	**30.48**	**3.75**	**5.29**	**3.69**	**5.22**

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

Main results. Tables 1–2 compare CSBOO with the baselines across ten forecasting tasks. CSBOO obtains the best result in 18 of 20 MAE/RMSE comparisons, and on PeMS08 lowers MAE/RMSE by 7.9%/4.0% relative to the strongest baseline. The advantage is consistent on highway flow and speed datasets with different graph sizes, and it also carries over to T-Drive and CHIBike inflow/outflow, whose spatial structure is grid based. This cross-regime consistency is important because the central challenge is not only to extract spatio-temporal features, but to determine when recent, daily, or weekly histories remain reliable for each node and horizon. The shared temporally causal field provides comparable candidate states without relying on a supplied adjacency, while mismatch-aware actions penalize unreliable periodic references. Markov transport then preserves cross-horizon orbit coherence, and the Feynman–Kac correction adapts transported mass to sample-

Table 3: Ablations of CSBOO reliability and orbit-inference mechanisms.

	PeMS08		PeMS07		PeMSD7(L)		T-Drive Avg.	
Variant	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
Full CSBOO	**12.78**	**22.39**	**18.99**	**32.23**	**2.86**	**5.77**	**17.11**	**30.35**
w/o Mismatch	13.15	22.92	19.31	32.71	2.94	5.91	17.37	30.78
w/o Markov Transport	12.98	22.67	19.17	32.48	2.91	5.85	17.25	30.60
w/o Responsibility Calibration	12.84	22.48	19.06	32.32	2.88	5.80	17.16	30.43
w/o Anchor Loss	12.91	22.57	19.11	32.42	2.90	5.83	17.21	30.53
w/o Path Inference (Softmax)	13.28	23.10	19.44	32.92	2.97	5.95	17.46	30.92

Table 4: Held-out diagnostics of learned orbit behavior.

Diagnostic	**Result**	**Diagnostic**	**Result**
Daily level mismatch–mass ρ	-0.34	Weekly level mismatch–mass ρ	-0.81
Daily slope mismatch–mass ρ	-0.21	Weekly slope mismatch–mass ρ	-0.73
Mass variation—Full	0.014	Mass variation—w/o Markov	0.028
Responsibility–mass KL—Full	0.35	Responsibility–mass KL—w/o Calibration	1.16
Mass–candidate error rank ρ	-0.42	Marginal entropy $H(\boldsymbol{\mu})$	0.77
Transition entropy $H(\mathbf{T})$	0.97		

and node-specific evidence. The results therefore show that the gains arise from reliability-aware sequential orbit inference rather than dataset-specific fusion or backbone capacity alone.

Ablation results. Tables 3–4 trace the gains to the three method challenges. Removing mismatch causes the largest drop among single-component removals, showing the importance of reliability adaptation when daily or weekly references become misleading; removing Markov transport breaks the intended coherence of orbit relevance across forecast horizons; and dropping responsibility calibration or anchor regularization weakens the predictive semantics and stability of orbit mass. Direct softmax fusion discards the sequential Feynman–Kac path structure and therefore performs worst. The orbit diagnostics show the complementary behavior prescribed by the method: mass is suppressed for mismatched or inaccurate candidates, evolves coherently across horizons, and aligns more closely with anchored responsibility under calibration. Together, the tests connect mismatch to reliability (C1), Markov inference to cross-horizon coherence (C2), and anchoring with calibration to predictive responsibility (C3), attributing the gains to sequential orbit inference rather than backbone capacity alone.

6 Conclusion

Reliable multi-period forecasting requires more than exposing a model to recent, daily, and weekly histories: it must identify which reference remains reliable at each node and horizon while keeping that relevance coherent and interpretable. We introduced CSBOO, which replaces independent multi-period fusion with sequential inference over the responsibilities of semantically anchored predictive orbits in a shared temporally causal field. Horizon-aware mismatch supplies reliability evidence, Markov transport propagates orbit belief across forecast steps, and the Feynman–Kac correction adapts transported mass to sample- and node-specific conditions; responsibility calibration gives the resulting marginals predictive meaning, and their barycenter forms the forecast. The path-space characterization and exact terminal-marginal recursion provide a theoretical foundation for this construction. Across ten tasks, CSBOO obtains the best result in 18 of 20 MAE/RMSE comparisons, including 7.9%/4.0% reductions on PeMS08 relative to the strongest baseline. Its consistent gains across highway and urban-grid regimes, together with the ablations and orbit diagnostics, attribute the improvement to reliability-aware sequential inference rather than unconstrained fusion.

Reproducibility statement

Appendix A documents the benchmark datasets and evaluation splits. Appendices B–L specify the causal input construction, path recursion, proofs, model components, paper-to-code mapping, and released hyperparameters. The supplementary code provides the corresponding model implementation and invariant checks.

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

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

Yu Chen, Wei Deng, Shikai Fang, Fengpei Li, Nicole Tianjiao Yang, Yikai Zhang, Kashif Rasul, Shandian Zhe, Anderson Schneider, and Yuriy Nevmyvaka. Provably convergent Schrödinger bridge with applications to probabilistic time series imputation. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 4485–4513. PMLR, 2023. URL https://proceedings.mlr.press/v202/chen23f.html.

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.

Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc., 2013. URL https://proceedings.neurips.cc/paper_files/paper/2013/file/af21d0c97db2e27e13572cbf59eb343d-Paper.pdf.

Pierre Del Moral. *Feynman–Kac Formulae: Genealogical and Interacting Particle Systems With Applications*. Springer, 2004. doi: 10.1007/978-1-4684-9393-1.

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.

Sarthak Jain and Byron C. Wallace. Attention is not Explanation. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics:*

Human Language Technologies, Volume 1 (Long and Short Papers), pp. 3543–3556. Association for Computational Linguistics, 2019. doi: 10.18653/v1/N19-1357. URL https://aclanthology.org/N19-1357/.

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.

Christian Léonard. A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete and Continuous Dynamical Systems*, 34(4):1533–1574, 2014. doi: 10.3934/dcds.2014.34.1533.

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.

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.

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.

Alexander Y. Tong, Nikolay Malkin, Kilian Fatras, Lazar Atanackovic, Yanlei Zhang, Guillaume Huguet, Guy Wolf, and Yoshua Bengio. Simulation-free Schrödinger bridges via score and flow matching. In *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pp. 1279–1287. PMLR, 2024. URL https://proceedings.mlr.press/v238/tong24a.html.

Gefei Wang, Yuling Jiao, Qian Xu, Yang Wang, and Can Yang. Deep generative learning via Schrödinger bridge. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 10794–10804. PMLR, 2021. URL https://proceedings.mlr.press/v139/wang21l.html.

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.

Huaxiu Yao, Xianfeng Tang, Hua Wei, Guanjie Zheng, and Zhenhui Li. Revisiting spatial-temporal similarity: A deep learning framework for traffic prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pp. 5668–5675, 2019. doi: 10.1609/aaai.v33i01.33015668.

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.

Junbo Zhang, Yu Zheng, and Dekang Qi. Deep spatio-temporal residual networks for citywide crowd flows prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31, 2017. doi: 10.1609/AAAI.V31I1.10735.

Weiguo Zhu, Yongqi Sun, Xintong Yi, Yan Wang, and Zhen Liu. A correlation information-based spatiotemporal network for traffic flow forecasting. *Neural Computing and Applications*, 35(28): 21181–21199, 2023. doi: 10.1007/s00521-023-08831-3.

A Dataset Details

Table 5: Benchmark datasets and evaluation splits.

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

B Causal Input Construction

Let $\mathcal{V}$ be a set of N spatial nodes and let $\mathbf{x}_s \in \mathbb{R}^N$ be the signal at original time index s. A sample ending at time t uses L historical steps to predict $\mathbf{Y}_{t+1:t+H} \in \mathbb{R}^{H\times N}$. If one day contains S sampling steps, the daily and weekly periods are $P_{\mathrm{D}} = S$ and $P_{\mathrm{W}} = 7S$. With $\ell = 1{:}L$ and $h = 1{:}H$, the temporal references are

$$\begin{aligned} r_{\ell,n} &= x_{t-L+\ell,n}, & d_{\ell,n} &= x_{t-L+\ell-P_{\mathrm{D}},n}, & w_{\ell,n} &= x_{t-L+\ell-P_{\mathrm{W}},n}, \\ d^{+}_{h,n} &= x_{t+h-P_{\mathrm{D}},n}, & w^{+}_{h,n} &= x_{t+h-P_{\mathrm{W}},n}. \end{aligned} \tag{21}$$

Let $I^{\mathrm{obs}}_{s,n} \in \{0,1\}$ indicate that $x_{s,n}$ is observed. Historical masks and horizon-anchor availability are

$$\begin{aligned} m^{\mathrm{D}}_{\ell,n} &= I^{\mathrm{obs}}_{t-L+\ell-P_{\mathrm{D}},n}, & m^{\mathrm{W}}_{\ell,n} &= I^{\mathrm{obs}}_{t-L+\ell-P_{\mathrm{W}},n}, \\ \eta^{\mathrm{D}}_{h,n} &= I^{\mathrm{obs}}_{t+h-P_{\mathrm{D}},n}\mathbb{I}[t+h-P_{\mathrm{D}} \le t], & \eta^{\mathrm{W}}_{h,n} &= I^{\mathrm{obs}}_{t+h-P_{\mathrm{W}},n}\mathbb{I}[t+h-P_{\mathrm{W}} \le t]. \end{aligned} \tag{22}$$

Unavailable values are zero-filled only after their masks are formed. Thus an anchor is usable only when it is observed and lies no later than t, making the construction strictly causal. We assume that $r_{L,n} = x_{t,n}$ is available for every prediction target; samples violating this condition are excluded during preprocessing. The model receives separate history and horizon axes,

$$\begin{aligned} X^{\mathrm{hist}}_{\ell,n} &= [r_{\ell,n}, \mathrm{tod}_{\ell,n}, \omega_{\ell,n}, d_{\ell,n}, m^{\mathrm{D}}_{\ell,n}, w_{\ell,n}, m^{\mathrm{W}}_{\ell,n}] \in \mathbb{R}^7, \\ A^{+}_{h,n} &= [d^{+}_{h,n}, \eta^{\mathrm{D}}_{h,n}, w^{+}_{h,n}, \eta^{\mathrm{W}}_{h,n}] \in \mathbb{R}^4. \end{aligned} \tag{23}$$

Here $\mathrm{tod} \in [0,1)$ is normalized time of day and $\omega \in \{0,\dots,6\}$ is day of week. Known future calendar variables are represented by $\Phi^{+} = [\boldsymbol{\phi}_1,\dots,\boldsymbol{\phi}_H]^{\top} \in \mathbb{R}^{H\times 4}$, with

$$\boldsymbol{\phi}_h = [\sin(2\pi \mathrm{tod}^{+}_h), \cos(2\pi \mathrm{tod}^{+}_h), \sin(2\pi \omega^{+}_h/7), \cos(2\pi \omega^{+}_h/7)].$$

The implementation computes these variables from the last timestamp, concatenates the learned horizon embedding $\boldsymbol{\xi}_h$, and uses $p_h = P_\phi([\boldsymbol{\phi}_h; \boldsymbol{\xi}_h])$. Hence the formal inputs are $(X^{\mathrm{hist}}, A^{+}, \Phi^{+})$; no equality between L and H is required.

C Gibbs Path Characterization and Exact Terminal-Marginal Recursion

Fix a node n and prefix horizon h, suppress the node index, and retain the convention $\mu_h(k) \equiv \mu^k_{h,n}$ and $a_h(k) \equiv a^k_{h,n}$. Define the path cost $C_h(k_{0:h}) = \sum_{s=1}^{h} a_s(k_s)$. Since the softmax initial law and transition matrices are strictly positive, the reference law in Eq. (16) has full support. For $\epsilon > 0$, let

$$Z_h = \mathbb{E}_{\mathbb{P}_h}\left[e^{-C_h/\epsilon}\right], \qquad \mathbb{Q}^*_h(k_{0:h}) = \frac{\mathbb{P}_h(k_{0:h})e^{-C_h(k_{0:h})/\epsilon}}{Z_h}. \tag{24}$$

For any path law $\mathbb{Q}$,

$$\mathbb{E}_{\mathbb{Q}}[C_h] + \epsilon\,\mathrm{KL}(\mathbb{Q}\|\mathbb{P}_h) = \epsilon\,\mathrm{KL}(\mathbb{Q}\|\mathbb{Q}^*_h) - \epsilon \log Z_h. \tag{25}$$

The second term is constant in $\mathbb{Q}$, while the KL term is nonnegative and vanishes only at $\mathbb{Q} = \mathbb{Q}^*_h$. This proves both the Gibbs form and uniqueness in Proposition 1.

To obtain the implemented recursion, define an unnormalized terminal forward message

$$\begin{aligned} \Gamma_0(i) &= \mu_0(i), \\ \Gamma_s(j) &= e^{-a_s(j)/\epsilon} \sum_{i \in \mathcal{O}} \Gamma_{s-1}(i)\mathbf{T}_s(i,j). \end{aligned} \tag{26}$$

Expanding Eq. (26) inductively sums the weight $\mu_0(k_0)\prod_{r=1}^{s}\mathbf{T}_r(k_{r-1},k_r)e^{-\sum_{r=1}^{s}a_r(k_r)/\epsilon}$ over all paths ending at j. Hence the normalized terminal message is exactly the terminal marginal of $\mathbb{Q}^*_s$:

$$\mu_s(j) = \frac{\Gamma_s(j)}{\sum_k \Gamma_s(k)}. \tag{27}$$

Writing $\Gamma_{s-1} = c_{s-1}\mu_{s-1}$ in Eq. (26), the positive scalar c_{s-1} cancels in Eq. (27). The remaining Markov prediction and normalized potential correction are precisely Eqs. (14)– (15).

D Normalization, Convexity, and Masking

The base case is $\boldsymbol{\mu}_{0,n} = \mathrm{softmax}(\bar{o}_{0,n}) \in \Delta^2$. Assume $\boldsymbol{\mu}_{h-1,n} \in \Delta^2$. Each $\mathbf{T}_h$ is row stochastic, so $\widetilde{\mu}^j_{h,n} = \sum_i \mu^i_{h-1,n}\mathbf{T}_h(i,j)$ is nonnegative and

$$\sum_j \widetilde{\mu}^j_{h,n} = \sum_i \mu^i_{h-1,n} \sum_j \mathbf{T}_h(i,j) = 1. \tag{28}$$

For finite actions and $\epsilon > 0$, every factor $e^{-a^j_{h,n}/\epsilon}$ is positive, so the denominator in Eq. (15) is strictly positive. The corrected entries are therefore nonnegative and sum to one. Induction gives $\boldsymbol{\mu}_{h,n} \in \Delta^2$ for all h, and Eq. (1) is a convex combination of the candidates.

Holding other actions finite and taking $a^j_{h,n} \to +\infty$ sends the j-th corrected numerator to zero, establishing exact theoretical removal in the limit. The implementation instead uses a large finite M; it therefore produces numerically negligible, not mathematically exact, unavailable-orbit mass. In the limiting mask interpretation, the forecast is a convex combination of the usable candidates only.

E Mismatch-to-Mass Monotonicity

Fix (h,n), $\epsilon > 0$, a finite action vector, and a transported prior satisfying $\widetilde{\mu}^i > 0$ for every i. Hold the learned emission and all other mismatch terms, and hence all actions except a^j, fixed. Write $w^i = \widetilde{\mu}^i e^{-a^i/\epsilon}$ and $Z = \sum_i w^i$, so $\mu^i = w^i/Z$. Direct differentiation gives

$$\frac{\partial \mu^j}{\partial a^j} = -\frac{1}{\epsilon}\mu^j(1-\mu^j) < 0, \qquad \frac{\partial \mu^k}{\partial a^j} = \frac{1}{\epsilon}\mu^k\mu^j > 0 \quad (k \neq j). \tag{29}$$

Because $\lambda^j_E, \lambda^j_S > 0$, the chain rule shows that increasing E^j or S^j strictly decreases the corresponding periodic mass whenever that mass is interior. The cross-period disagreement I^{DW} enters both

periodic actions with coefficient $\lambda_I > 0$. Let $\mu^{\mathrm{per}} = \mu^{\mathrm{D}} + \mu^{\mathrm{W}} = 1 - \mu^{\mathrm{R}}$. Increasing I^{DW} shifts both periodic actions equally, and

$$\frac{\partial \mu^{\mathrm{per}}}{\partial I^{\mathrm{DW}}} = -\frac{\lambda_I}{\epsilon}\,\mu^{\mathrm{per}}(1-\mu^{\mathrm{per}}) = -\frac{\lambda_I}{\epsilon}\,(\mu^{\mathrm{D}}+\mu^{\mathrm{W}})\mu^{\mathrm{R}} < 0. \tag{30}$$

Thus increasing I^{DW} suppresses total periodic mass relative to the recent orbit.

F Gauge Invariance and Canonical Parameterization

For any $z \in \mathbb{R}^K$ and scalar c,

$$\mathrm{softmax}(z + c\mathbf{1})_j = \frac{e^{z_j+c}}{\sum_k e^{z_k+c}} = \frac{e^{z_j}}{\sum_k e^{z_k}} = \mathrm{softmax}(z)_j. \tag{31}$$

Subtracting the recent emission score therefore selects one representative from an additive equivalence class without changing the emission probabilities. Likewise, subtracting the mean of each transition-score row selects a canonical row representative without changing its softmax transition probabilities. These operations remove redundant additive parameterizations and provide canonical logits. The predictive semantics of the resulting orbit mass are supplied separately by the semantic-anchor and responsibility-calibration mechanisms.

G Responsibility Calibration and Anchor Semantics

By the preprocessing assumption in Appendix B, the final recent observation is available, so $\chi^{\mathrm{R}}_{h,n} = 1$ and the denominator in Eq. (19) is positive. For $\tau_{\mathrm{cal}} > 0$, every usable exponential is finite and strictly positive; normalization therefore defines a unique $\boldsymbol{\pi}$ on the available support. For fixed candidates and target, Gibbs' inequality gives $\mathrm{KL}(\boldsymbol{\pi}\|\boldsymbol{\mu}) \geq 0$ over the probability simplex on the available support, with equality exactly when $\boldsymbol{\mu} = \boldsymbol{\pi}$. This proves Proposition 4 under the prescribed anchor-based predictive-responsibility rule.

The anchors supply the recent/daily/weekly labels by construction, while $|\Delta^k_{h,n}| < R^{\mathrm{def}}_{t,n}$ bounds candidate drift. Candidate k therefore lies in $[\bar{c}^k_{h,n} - R^{\mathrm{def}}_{t,n}, \bar{c}^k_{h,n} + R^{\mathrm{def}}_{t,n}]$. If $|\bar{c}^i_{h,n} - \bar{c}^j_{h,n}| > 2R^{\mathrm{def}}_{t,n}$, the two admissible intervals are disjoint. When anchors coincide or their intervals overlap, boundedness alone cannot distinguish their predictive responsibilities; the target-conditioned $\boldsymbol{\pi}$ and $\mathcal{L}_{\mathrm{cal}}$ provide the responsibility-calibration criterion during training.

H Backbone and Decoder Details

This appendix expands the modules summarized in Section 4.1. For $x \in \{r, d, w\}$, define the first and second temporal differences by

$$\nabla x_{1,n} = 0, \qquad \nabla x_{\ell,n} = x_{\ell,n} - x_{\ell-1,n} \quad (\ell > 1),$$
$$\nabla^2 x_{1,n} = 0, \qquad \nabla^2 x_{\ell,n} = \nabla x_{\ell,n} - \nabla x_{\ell-1,n} \quad (\ell > 1).$$

Unavailable boundary differences are set to zero. The feature lift at each history position is

$$\begin{aligned}\mathbf{f} = [&r, \nabla r, \nabla^2 r, m^{\mathrm{D}} d, m^{\mathrm{D}}(r-d), \widetilde{m}^{\mathrm{D}}\nabla d, m^{\mathrm{W}} w, m^{\mathrm{W}}(r-w), \widetilde{m}^{\mathrm{W}}\nabla w,\\ &\sin(2\pi\mathrm{tod}), \cos(2\pi\mathrm{tod}), \sin(2\pi\omega/7), \cos(2\pi\omega/7), m^{\mathrm{D}}, m^{\mathrm{W}}],\end{aligned} \tag{32}$$

where $\widetilde{m}^k_\ell = m^k_\ell m^k_{\ell-1}$ for $\ell > 1$ and $\widetilde{m}^k_1 = 0$. Invalid periodic values, residuals, and slopes are zeroed before projection. With $U^{(0)}_{\ell,n} = W_{\mathrm{in}} f_{\ell,n} + v_n$, a causal block of width κ and dilation d_{dil} is

$$\begin{aligned}[A_{d_{\mathrm{dil}}}, G_{d_{\mathrm{dil}}}] &= \mathrm{Conv}_{\kappa,d_{\mathrm{dil}}}(\mathrm{PadLeft}_{(\kappa-1)d_{\mathrm{dil}}}(\mathrm{LN}(U))),\\ \mathcal{T}_{d_{\mathrm{dil}}}(U) &= U + \mathrm{Drop}\left(\mathrm{Conv}_1[\tanh(A_{d_{\mathrm{dil}}}) \odot \sigma(G_{d_{\mathrm{dil}}})]\right).\end{aligned} \tag{33}$$

After stacking the blocks, define $\mathbf{h}^{\mathrm{raw}}_n = [r_{1:L,n}, \nabla r_{1:L,n}, m^{\mathrm{D}}(r-d)_{1:L,n}, m^{\mathrm{W}}(r-w)_{1:L,n}] \in \mathbb{R}^{4L}$. The pre-spatial state is

$$s^{(0)}_n = \mathrm{LN}\left(U_{L,n} + \frac{1}{L}\sum_{\ell=1}^{L} U_{\ell,n} + W_{\mathrm{raw}}\mathbf{h}^{\mathrm{raw}}_n\right). \tag{34}$$

For the spatial kernels in Eq. (3), forward and reverse messages are combined as

$$\begin{aligned} M_i^{m,+} &= \sum_j F_{ij}^m V_j^m, \qquad M_i^{m,-} = \sum_j R_{ij}^m V_j^m, \\ \mathsf{Adv}_i &= W_A \,\mathrm{cat}_m \,\tfrac{1}{2}(M_i^{m,+} - M_i^{m,-}), \\ \mathsf{Dif}_i &= W_D \,\mathrm{cat}_m \left[\tfrac{1}{2}(M_i^{m,+} + M_i^{m,-}) - V_i^m\right], \\ \mathcal{K}_{\mathrm{sp}}(X)_i &= \mathrm{LN}\left(X_i + \sigma(W_g T_i) \odot \Psi(T_i)\right), \quad T_i = [\bar{X}_i, \mathsf{Adv}_i, \mathsf{Dif}_i]. \end{aligned} \tag{35}$$

Here Ψ is a linear–GLU–dropout–linear update. The odd branch carries directed advection; the mean-relative branch supplies a diffusion-like message.

For the phase decoder, use the same projected phase $p_h = P_\phi(\psi_h)$ and split $z_{h,n}$ into two-dimensional modes. Each mode uses a horizon-conditioned base angle and contraction,

$$\begin{aligned} \vartheta_{h,m}^{(1)} &= \vartheta_{\max} \tanh(\theta_m + [W_\theta p_h]_m), \\ \upsilon_{h,m}^{(1)} &= \upsilon_{\min} + (\upsilon_{\max} - \upsilon_{\min})\sigma(\beta_m + [W_r p_h]_m), \\ \bar{z}_{h,n,m} &= (\upsilon_{h,m}^{(1)})^h \,\mathrm{Rot}(h\vartheta_{h,m}^{(1)}) z_{h,n,m}, \end{aligned} \tag{36}$$

The factors $(\upsilon_{h,m}^{(1)})^h$ and $h\vartheta_{h,m}^{(1)}$ implement the corresponding h-step contraction and accumulated rotation for mode m. Here m indexes the two-dimensional latent modes, θ_m is a learnable base rotation parameter, β_m is a learnable base contraction logit, and

$$\mathrm{Rot}(\vartheta) = \begin{bmatrix} \cos\vartheta & -\sin\vartheta \\ \sin\vartheta & \cos\vartheta \end{bmatrix}.$$

Here $0 < \upsilon_{\min} < \upsilon_{\max} < 1$. The exact state returned to the heads is

$$q_{h,n} = \mathrm{LN}\left(\bar{z}_{h,n} + z_{h,n} + \Psi_q([\bar{z}_{h,n}, z_{h,n}, p_h, \upsilon_n])\right). \tag{37}$$

I Exact Mismatch, Candidates, and Numerical Recursion

The compact operator in Eq. (6) expands to separate valid attention masses and statistics,

$$Z_{h,n}^{k,0} = \sum_{\ell=1}^{L} \alpha_{h,n,\ell} m_{\ell,n}^k, \qquad Z_{h,n}^{k,1} = \sum_{\ell=1}^{L} \alpha_{h,n,\ell} \widetilde{m}_{\ell,n}^k,$$

$$E_{h,n}^k = \sqrt{\frac{\sum_{\ell=1}^{L} \alpha_{h,n,\ell} m_{\ell,n}^k (r_{\ell,n} - b_{\ell,n}^k)^2}{\max(Z_{h,n}^{k,0}, \delta)} + \delta}, \tag{38}$$

$$S_{h,n}^k = \frac{\sum_{\ell=1}^{L} \alpha_{h,n,\ell} \widetilde{m}_{\ell,n}^k |\nabla r_{\ell,n} - \nabla b_{\ell,n}^k|}{\max(Z_{h,n}^{k,1}, \delta)}. \tag{39}$$

For cross-period disagreement, let $Z_{h,n}^{\mathrm{DW}} = \sum_\ell \alpha_{h,n,\ell} m_{\ell,n}^{\mathrm{D}} m_{\ell,n}^{\mathrm{W}}$ and

$$I_{h,n}^{\mathrm{DW}} = \sqrt{\frac{\sum_{\ell=1}^{L} \alpha_{h,n,\ell} m_{\ell,n}^{\mathrm{D}} m_{\ell,n}^{\mathrm{W}} (d_{\ell,n} - w_{\ell,n})^2}{\max(Z_{h,n}^{\mathrm{DW}}, \delta)} + \delta}. \tag{40}$$

A periodic orbit is usable only if its horizon anchor is available and $Z_{h,n}^{k,0} > \delta$. When level support exists but no adjacent valid pair exists, its slope term is defined as zero.

Let $s_{t,n}^{\mathrm{loc}} = \max\{[(L^{-1}\sum_\ell (\nabla r_{\ell,n})^2) + \delta]^{1/2}, s_{\min}\}$ and $R_{t,n}^{\mathrm{def}} = \gamma s_{t,n}^{\mathrm{loc}}$. The supplied heads inside Eq. (8) are

$$\begin{aligned} \widetilde{\Delta}_{h,n}^{\mathrm{R}} &= [W_s s_n]_h + w_z^\top z_{h,n} + w_q^\top q_{h,n}, \\ \widetilde{\Delta}_{h,n}^{\mathrm{D}} &= w_d^\top q_{h,n}, \qquad \widetilde{\Delta}_{h,n}^{\mathrm{W}} = w_w^\top q_{h,n}, \qquad \Delta_{h,n}^k = R_{t,n}^{\mathrm{def}} \tanh(\widetilde{\Delta}_{h,n}^k). \end{aligned} \tag{41}$$

Their readouts are zero-initialized, so training starts from persistence, daily recurrence, and weekly recurrence.

The learned temperature is bounded by

$$\epsilon = \epsilon_{\min} + (\epsilon_{\max} - \epsilon_{\min})\sigma(\theta_\epsilon), \qquad 0 < \epsilon_{\min} < \epsilon_{\max}. \tag{42}$$

For an indexed collection (x_i), define $\mathrm{LSE}_i(x_i) := \log \sum_i \exp(x_i)$. The implementation evaluates Eqs. (14)–(15) directly in log mass as

$$\begin{aligned} \log \widetilde{\mu}^j_{h,n} &= \mathrm{LSE}_{i\in\mathcal{O}}(\log \mu^i_{h-1,n} + \log \mathbf{T}_h(i,j)), \\ \log \mu^j_{h,n} &= \log \widetilde{\mu}^j_{h,n} - a^j_{h,n}/\epsilon - \mathrm{LSE}_{k\in\mathcal{O}}(\log \widetilde{\mu}^k_{h,n} - a^k_{h,n}/\epsilon), \end{aligned} \tag{43}$$

which prevents underflow while preserving exactly the normalized probability recursion in the main paper.

J Implementation-Level Forward Pass

For each batch, the preferred interface receives X^{hist} on the L axis and A^+ on the H axis. For compactness, node indices are suppressed in the following algorithmic summary. The implemented computation is:

1. Run the causal backbone, retarded field, shared horizon-refinement spatial layer, and decoder to obtain s, $\alpha_{1:H}$, $u_{1:H}$, $z_{1:H}$, and $q_{1:H}$.
2. Compute $E^k_h, S^k_h, I^{\mathrm{DW}}_h$ with valid attention-mass denominators, and construct the bounded anchored candidates.
3. Set the gauge-fixed $\log \boldsymbol{\mu}_0 = \log \mathrm{softmax}(\bar{o}_0)$ and compute ϵ from Eq. (42).
4. For $h = 1, \ldots, H$: compute the gauge-fixed emission and action; mask unavailable periodic states; row-center B_h and compute $\mathbf{T}_h$; apply Eqs. (14)–(15); and append $\widehat{y}_h = \sum_k \mu^k_h c^k_h$.
5. During training, form $\boldsymbol{\pi}_h$ and evaluate Eq. (20); omit target-dependent terms at inference.
6. Stack the H forecasts and retain detached measures and transitions for diagnostics.

K Paper-to-Code Correspondence

Table 6: Correspondence to the supplied PyTorch implementation.

Paper object	Source object
$X^{\mathrm{hist}}, A^+, \Phi^+$	`x`, `horizon_anchors`, and `future_phase(x)`
$s, \psi_h, \alpha_h, u_h, z_h, q_h$	`state`, `phases`, `delay_weights`, local `delayed`, `horizon_states`, and `flow_states`
$E^{\mathrm{D}}_h, E^{\mathrm{W}}_h, S^{\mathrm{D}}_h, S^{\mathrm{W}}_h, I^{\mathrm{DW}}_h$	`daily_energy`, `weekly_energy`, `daily_slope`, `weekly_slope`, and `orbit_disagreement`
$\Delta^k, c^{\mathrm{R}}, c^{\mathrm{D}}, c^{\mathrm{W}}$	`deformations` and `candidates[..., 0:3]`
$o_0, \boldsymbol{\mu}_0$	`initial_measure(state)` after recent-orbit gauge fixing
e_h	`emission(cat(delayed, phase, node))`, where the local `delayed` variable is `horizon_states[:, h]`
a_h	`action`
B_h and $\mathbf{T}_h$	transition-score matrix: `transition_action[h]`; transition matrix: `transition`
ϵ	`entropic_temperature_raw` after bounded sigmoid
$\log \boldsymbol{\mu}_h, \boldsymbol{\mu}_h$	`log_measure`, `measure`
$\widehat{y}_h$	`(measure * candidates).sum(-1)`
$\boldsymbol{\pi}_h, \mathcal{L}_{\mathrm{cal}}, \mathcal{L}_{\mathrm{anc}}$	`responsibility_calibration_loss(target)`

L Released Implementation Configuration

Table 7 isolates code-level choices from the method definition. They describe the supplied implementation and can be tuned without changing the operator in Sections 4.1–4.5.

Table 7: Released implementation defaults.

Component	Released setting
History / forecast length	$L = 12$, $H = 12$
Input interface	seven-channel L-axis history plus four-channel H-axis anchors; legacy nine-channel packing supported when $L \geq H$
Backbone width / heads / dropout	$d = 128$, $A = 4$, dropout $= 0.10$
Temporal stack	kernel width 3, dilations $(1, 2, 4)$
Spatial temperature multiplier	learned in $(0.75, 1.25)$
Retarded field	rank 24; temperature in $(0.75, 1.25)$; recency-bias slope -0.15
Rotation / contraction bounds	horizon embedding width 4; $\vartheta_{\max} = 0.20\pi$; $(\upsilon_{\min}, \upsilon_{\max}) = (0.94, 0.999)$
Numerical and deformation constants	$\delta = 10^{-6}$, $s_{\min} = 10^{-3}$, $\gamma = 0.5$, unavailable-action value $M = 10^4$
Shared entropic temperature	$(\epsilon_{\min}, \epsilon_{\max}) = (0.70, 1.30)$
Initialization	initial-measure and emission output maps zero-initialized; transition scores identity-initialized
Training objective	MAE prediction loss; $\tau_{\text{cal}} = 0.5$, $\lambda_{\text{cal}} = 1$, and $\lambda_{\text{anc}} = 0.05$