MORE SENSORS ONLY ONE FIELD: RETHINKING CONTINUAL SPATIO-TEMPORAL FORECASTING

Anonymous authors
Paper under double-blind review

ABSTRACT

Continual spatio-temporal forecasting supports traffic management and environmental monitoring under evolving dynamics and expanding sensor networks. However, conventional graph-based continual learning methods tie forecasting representations to the current sensor layout, so adding sensors can alter how previously learned spatial relationships are represented. ***Our key insight is that sensor expansion changes the evidence available about a process without necessarily changing the dynamics to be learned.*** Motivated by this insight, we propose **STFO** (Spatio-Temporal Field Operator), which parameterizes forecasting knowledge as a shared field-evolution operator while handling changing sensor layouts through observation and query interfaces. **STFO** lifts irregular sensor histories onto a fixed latent grid through normalized coordinate-based aggregation. This common computational domain enables reuse of learned spatial maps across observation sets without sensor-specific parameters. To accommodate process drift within this shared representation, a spectral descriptor summarizes variation across spatial scales and conditions Fourier propagation and attention, enabling the operator's responses to adjust to the current spatial regime. Coordinate-based decoding reads the evolved field at observed sensor locations and combines spatial corrections with local-history predictions. Experiments on PEMS-Stream, CA-Stream, and AIR-Stream demonstrate state-of-the-art average forecasting performance. Specifically, **STFO**-Large reduces average MAE relative to DOL by 8.4% on PEMS-Stream and 4.7% on CA-Stream. Our code is available at `https://anonymous.4open.science/r/STFO2027/`.

1 INTRODUCTION

Continual spatio-temporal forecasting is essential for traffic management and environmental monitoring systems that operate over extended periods. These systems must maintain accurate predictions as both underlying dynamics and sensor layouts evolve: seasonal and long-term changes alter the process being forecast, while newly deployed sensors expand observation coverage (Chen et al., 2021; Francies et al., 2025). This calls for reusing forecasting knowledge from earlier periods while incorporating new sensors and adapting to changing dynamics.

Existing spatio-temporal forecasting methods typically represent sensors as graph nodes and learn spatial dependencies through graph convolution or attention (Yu et al., 2018; Wu et al., 2019; Guo et al., 2019; Zheng et al., 2020). Continual methods extend these representations through experience replay, knowledge consolidation, and pattern memory (Chen et al., 2021; Wang et al., 2023a;b), or use online adaptation to separate stable and dynamic components (Wang et al., 2024; 2025).

Despite this progress, transferring forecasting knowledge across sensor layouts remains challenging, because ***changes in the underlying dynamics and observation layouts*** enter the same graph-based representation. In graph-based formulations, the spatial computation is defined directly over the current sensor set. Adding sensors can therefore change graph neighborhoods and aggregation weights, even when the underlying spatial dynamics among previously observed locations remain similar. Figure 1(a) quantifies this effect: on CA-Stream, where the sensor count grows from 480 to 1698, up to 72% of kNN edges are reconfigured after a single expansion. ***Our key insight is that such large changes in observation layouts need not imply comparable changes in underlying spatial relationships.*** Figure 1(b) provides supporting

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.

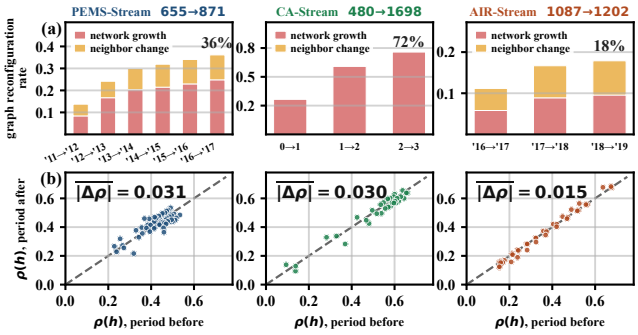

Figure 1: Motivation. (a) Fraction of kNN edges reconfigured after each expansion. (b) Pairwise spatial correlation.

evidence from the three streams. For each pair of consecutive periods, we compute the mean correlation $\rho(h)$ between sensor pairs at separation h and plot the values before and after expansion on the x and y axes, respectively. Points near the diagonal indicate similar correlations across periods. We first remove the shared time-of-day cycle to reduce its influence on the measured correlations and use a fixed sensor set shared across all periods to exclude changes in sensor composition. Across the three datasets, the average shift never exceeds 0.031, even where the sensor count grows by a factor of 3.5. These results suggest that sensor expansion changes how densely the underlying process is sampled, without necessarily changing the spatial dependencies to be learned.

The contrast between panels (a) and (b) motivates a separation that graph representations do not naturally provide: ***decouple knowledge of underlying spatial dynamics from how it interfaces with a particular observation layout.*** Existing methods typically adapt spatial representations defined over the current sensor set, coupling layout changes with adaptation to underlying dynamics. Ideally, when the layout changes while spatial dynamics remain similar, previously learned forecasting knowledge should remain reusable across layouts. The function-space perspective of neural operators (Li et al., 2020; Lu et al., 2021) offers a natural basis for this separation: sensor measurements provide observations of a spatio-temporal field at given coordinates, while a shared operator models its evolution. New sensors then contribute additional observations without requiring changes to the operator's parameterization. However, separating sensor layouts from underlying dynamics does not eliminate the need to adapt to ***drift in spatial dynamics***. The balance between broad spatial trends and localized variation can shift across periods, requiring the model to adjust its spatial responses. The resulting design principle is to ***maintain a fixed computational domain for knowledge reuse while adapting field evolution to changing spatial dynamics***.

Motivated by this principle, we introduce **STFO**, the ***Spatio-Temporal Field Operator***, which realizes the separation above through two complementary mechanisms. ❶ ***Sensor-independent field representation.*** **STFO** maps observations from changing sensor layouts onto a fixed latent grid through coordinate-based aggregation, representing forecasting knowledge without sensor-specific parameters. A coordinate-based decoder queries the evolved field at sensor locations and combines the resulting features with local histories to produce forecasts. The latent grid and parameter shapes remain fixed during fine-tuning across periods. ❷ ***Drift-adaptive field evolution.*** To accommodate changes in the underlying spatial dynamics, a compact spectral descriptor summarizes how variation is distributed across spatial scales and conditions the operator's channels on the current regime. A dual-path operator combines Fourier propagation for broad spatial structure with state-dependent attention for localized variation, enabling field evolution to adapt to changing dynamics.

Our contributions are threefold: ❶ We reveal a previously underexplored mismatch in continual spatio-temporal forecasting between changes in observation layout and changes in the underlying spatial dynamics, and propose **STFO** to separate them through sensor-independent field representation and drift-adaptive field evolution. ❷ For fixed neural maps and coordinates, we establish sensor-sampling consistency under explicit sampling assumptions and latent-grid consistency for a fixed spectral design. We separately quantify the effect of coordinate reassignment when the sensor set changes; ❸ We validate our method on three real-world datasets, demonstrating substantial improvements in forecasting accuracy over conventional graph-based continual learning methods.

2 Problem Formulation

Continual forecasting proceeds through periods $p = 1, \ldots, P$, with nested sensor sets $\mathcal{V}_p \subseteq \mathcal{V}_{p+1}$ and $N_p = |\mathcal{V}_p|$. Each period supplies observations from the current sensor set and updates the parameters learned in the previous period. Table 3 in Appendix B lists the notation in this paper.

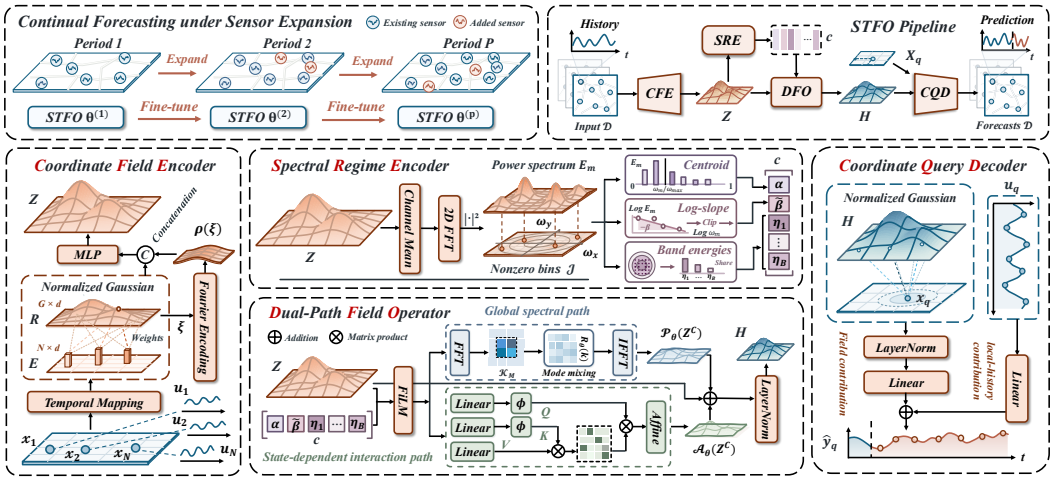

Figure 2: Overview of **STFO**. The upper panels illustrate continual forecasting under sensor expansion and the overall pipeline; the lower panels detail CFE, SRE, DFO, and CQD.

Coordinates and observations. Sensor i is assigned a coordinate $\boldsymbol{x}_i^{(p)} = \chi_p(\boldsymbol{s}_i) \in \Omega = [0,1]^2$, where $\boldsymbol{s}_i$ contains static descriptors such as geographic location and available attributes. The mapping χ_p depends on the current sensor set and is used consistently for encoding and querying within each period. Appendix D details the coordinate assignment procedure. Omitting p below, let $u_t : \Omega \to \mathbb{R}$ denote the spatial field and $y_{i,t} = u_t(\boldsymbol{x}_i) + \varepsilon_{i,t}$ its noisy observation. The input set is $\mathcal{D} = \{(\boldsymbol{x}_i, \boldsymbol{u}_i)\}_{i=1}^N$, where $\boldsymbol{u}_i = [y_{i,1}, \ldots, y_{i,T_h}]^\top$ is sensor i's observation history. For an observed query sensor q with history $\boldsymbol{u}_q$, the forecasting operator produces $\hat{\boldsymbol{y}}_q = \mathcal{G}_\theta(\mathcal{D})(\boldsymbol{x}_q) \in \mathbb{R}^{T_f}$, whose k-th component estimates $u_{T_h+k}(\boldsymbol{x}_q)$.

Discretization consistency. For a fixed coordinate representation, fixed neural maps, and a noiseless field history, consider sensor layouts $\mathcal{S} \subset \Omega$ with a fill distance $h_{\mathcal{S}} = \sup_{\boldsymbol{x}\in\Omega} \min_{\boldsymbol{x}_i\in\mathcal{S}} \|\boldsymbol{x} - \boldsymbol{x}_i\|_2$. Consistency requires predictions to converge to a common reference under an admissible sampling refinement: $\left\|\mathcal{G}_\theta(\mathcal{D}_{\mathcal{S}})(\boldsymbol{x}_q) - \overline{\mathcal{G}}_\theta(\boldsymbol{x}_q)\right\|_2 \longrightarrow 0$. Theorem 1 gives sampling bounds for fixed and shrinking bandwidths; Corollary 1 compares layouts sharing the same reference, and Theorem 2 bounds the error under latent-grid refinement. Appendix C provides the proofs and separately analyzes coordinate reassignment.

3 Methodology

STFO transfers forecasting knowledge through a shared latent field, as illustrated in Figure 2. The pipeline comprises four stages: CFE aggregates observations into spatial features; SRE summarizes their spectral regime; DFO evolves the conditioned field; and CQD decodes forecasts at sensor locations. We use a fixed Cartesian grid $\Xi_{h_g} = \{\boldsymbol{\xi}_g\}_{g=1}^G$ with n_g points per axis, spacing $h_g = 1/(n_g - 1)$, and $G = n_g^2$. Each period uses a $G \times d$ latent field, where d is the hidden width. Sensor expansion changes the input observation set, while all temporal, spatial, and output maps remain shared across sensors.

3.1 Coordinate Field Encoder

The encoder maps observations from irregular sensor layouts onto a fixed latent grid. A shared temporal map $\tau_\theta : \mathbb{R}^{T_h} \to \mathbb{R}^d$ encodes each sensor's history as $\boldsymbol{e}_i = \tau_\theta(\boldsymbol{u}_i)$. To lift these codes, define the Gaussian kernel $\kappa_\sigma(\boldsymbol{x}, \boldsymbol{x}') = \exp[-\|\boldsymbol{x} - \boldsymbol{x}'\|_2^2/(2\sigma^2)]$. At each grid location, the normalized weighted sum is

$$\boldsymbol{r}_g = \sum_{i=1}^N a_{gi}\boldsymbol{e}_i, \qquad a_{gi} = \frac{\kappa_{\sigma_{\text{enc}}}(\boldsymbol{\xi}_g, \boldsymbol{x}_i)}{\sum_{j=1}^N \kappa_{\sigma_{\text{enc}}}(\boldsymbol{\xi}_g, \boldsymbol{x}_j)}. \tag{1}$$

The learned bandwidth $\sigma_{\text{enc}} > 0$ controls the spatial extent of aggregation. The normalized weights give greater emphasis to nearby sensors and keep the lift a weighted average as the sensor count

changes. The aggregation is invariant to sensor ordering. Lemma 1 bounds the resulting lift error by the distance between sampling measures at a fixed bandwidth, and Lemma 2 shows that the error decays linearly in the fill distance when the bandwidth shrinks with it.

The field incorporates position information through $\boldsymbol{z}_g = \psi_\theta([\boldsymbol{r}_g \,\|\, \rho(\boldsymbol{\xi}_g)])$, where ρ is a fixed Fourier coordinate encoding (Tancik et al., 2020) and $\psi_\theta : \mathbb{R}^{d+d_\rho} \to \mathbb{R}^d$ is shared across the grid. Position features allow grid locations with similar aggregated features to receive different latent representations. In matrix form, the lift is $\mathbf{R} = \mathbf{AE}$, with $[\mathbf{A}]_{gi} = a_{gi}$ and $[\mathbf{E}]_{i,:} = \boldsymbol{e}_i^\top$. Since $\mathbf{A}$ has nonnegative entries and each row sums to one, each $\boldsymbol{r}_g$ is a convex combination of sensor codes, and the induced matrix norm satisfies $\|\mathbf{A}\|_\infty = 1$ for any sensor count. Stacking the resulting representations gives $\mathbf{Z} = [\boldsymbol{z}_1, \ldots, \boldsymbol{z}_G]^\top \in \mathbb{R}^{G\times d}$. Thus, the sensor count determines the input dimension of aggregation, while the latent forecasting state retains a fixed shape $G \times d$.

3.2 Spectral Regime Encoder

The spectral regime describes how variation is distributed over spatial scales. Broad spatial patterns concentrate energy at low frequencies, while fine-scale variation contributes to higher frequencies. SRE captures this distribution by averaging latent channels, $\bar{z}_g = d^{-1}\mathbf{1}_d^\top \boldsymbol{z}_g$, and transforming the resulting scalar grid $\bar{z} = \{\bar{z}_{\boldsymbol{j}}\}$, where $\boldsymbol{j} \in \{0, \ldots, n_g - 1\}^2$ indexes grid entries. We use a forward-normalized discrete Fourier transform with the corresponding inverse:

$$\widehat{\bar{z}}(\boldsymbol{k}) = \frac{1}{G}\sum_{\boldsymbol{j}} \bar{z}_{\boldsymbol{j}} e^{-2\pi \mathrm{i}\boldsymbol{k}\cdot\boldsymbol{j}/n_g}, \qquad \bar{z}_{\boldsymbol{j}} = \sum_{\boldsymbol{k}} \widehat{\bar{z}}(\boldsymbol{k}) e^{2\pi \mathrm{i}\boldsymbol{k}\cdot\boldsymbol{j}/n_g}. \tag{2}$$

Both transforms use phase nodes $\boldsymbol{j}/n_g$, whereas the latent-grid coordinates are $\boldsymbol{j}/(n_g - 1)$. Appendix C.5 accounts for this difference in the grid-consistency proof.

For a set $\mathcal{J}$ of nonzero Fourier bins with wavevectors $\boldsymbol{\omega}_m$, define $E_m = |\widehat{\bar{z}}(\boldsymbol{\omega}_m)|^2 + \epsilon_s$ and $\omega_m = \|\boldsymbol{\omega}_m\|_2 + \epsilon_f$, where $\epsilon_s, \epsilon_f > 0$ are numerical stabilizers. Let $\lambda_m = \log \omega_m$, $\mu_m = \log E_m$, and use an overbar to denote the average over $\mathcal{J}$, e.g., $\bar{\lambda} = |\mathcal{J}|^{-1}\sum_{m\in\mathcal{J}} \lambda_m$. We partition $\mathcal{J}$ into B logarithmically spaced radial frequency bands $\mathcal{B}_b$. Three complementary summaries form the regime descriptor:

$$\begin{aligned} \alpha &= \frac{\sum_{m\in\mathcal{J}} E_m \omega_m}{\omega_{\max}\sum_{m\in\mathcal{J}} E_m + \epsilon_s}, & \eta_b &= \frac{\sum_{m\in\mathcal{B}_b} E_m}{\sum_{m\in\mathcal{J}} E_m + \epsilon_s}, \\ \beta &= -\frac{\sum_{m\in\mathcal{J}}(\lambda_m - \bar{\lambda})(\mu_m - \bar{\mu})}{\sum_{m\in\mathcal{J}}(\lambda_m - \bar{\lambda})^2 + \epsilon_s}, & \boldsymbol{c} &= [\alpha, \tilde{\beta}, \eta_1, \ldots, \eta_B]^\top \in \mathbb{R}^{B+2}. \end{aligned} \tag{3}$$

Here $\omega_{\max} = \max_{m\in\mathcal{J}} \omega_m$ and $\tilde{\beta} = \mathrm{clip}(\beta, -10, 10)$. The centroid α summarizes the relative frequency location of energy; β is a regularized negative log–log slope describing radial spectral decay; and the band ratios η_b, $b = 1, \ldots, B$, capture energy contributions across spatial scales. The positive stabilizers ensure continuity under latent-field perturbations, while Theorem 3 quantifies stability under spectral-power perturbations. For the grid-consistency analysis, we keep $\mathcal{J}$ and its band partition fixed across grid resolutions.

3.3 Dual-Path Field Operator

Each operator block uses the spectral descriptor $\boldsymbol{c}$ to modulate latent channels before spatial propagation. A FiLM-style map (Perez et al., 2018) produces $(\boldsymbol{\gamma}, \boldsymbol{\delta}) = 0.1\tanh(\mathrm{MLP}_\theta(\boldsymbol{c})) \in \mathbb{R}^d \times \mathbb{R}^d$. The conditioned field is $\mathbf{Z}^{\mathrm{c}} = \mathbf{Z} \odot \big(\mathbf{1}_G(\mathbf{1}_d + \boldsymbol{\gamma})^\top\big) + \mathbf{1}_G\boldsymbol{\delta}^\top$, where $\odot$ denotes elementwise multiplication. Equivalently, each grid feature is updated as $\boldsymbol{z}_g^{\mathrm{c}} = \boldsymbol{z}_g \odot (\mathbf{1}_d + \boldsymbol{\gamma}) + \boldsymbol{\delta}$, using the same channel gains and offsets at all grid locations. The bounded gain and offset preserve a direct connection to the original field while adapting its channel emphasis. Stacked blocks have separate parameters and share the descriptor extracted from the initial field.

Global spectral propagation. Following Fourier neural operators (Li et al., 2020), the Fourier path communicates through learned spatial modes. For a retained low-frequency set $\mathcal{K}_M$, its update is

$$\mathcal{P}_\theta(\mathbf{Z}^{\mathrm{c}}) = \mathcal{F}_{2D}^{-1}[\mathbf{1}\{\boldsymbol{k} \in \mathcal{K}_M\}\, \mathbf{R}_\theta(\boldsymbol{k})\, \mathcal{F}_{2D}[\mathbf{Z}^{\mathrm{c}}](\boldsymbol{k})]\,. \tag{4}$$

Equivalently, at grid location g,

$$\big(\mathcal{P}_\theta(\mathbf{Z}^{\mathrm{c}})\big)_g = \sum_{\boldsymbol{k}\in\mathcal{K}_M} \mathbf{R}_\theta(\boldsymbol{k})\,\widehat{\mathbf{Z}}^{\mathrm{c}}(\boldsymbol{k})\,e^{2\pi \mathrm{i}\boldsymbol{k}\cdot\boldsymbol{t}_g}, \qquad \widehat{\mathbf{Z}}^{\mathrm{c}}(\boldsymbol{k}) = \frac{1}{G}\sum_{g=1}^{G} \boldsymbol{z}_g^{\mathrm{c}} e^{-2\pi \mathrm{i}\boldsymbol{k}\cdot\boldsymbol{t}_g}, \tag{5}$$

where $\boldsymbol{t}_g$ is the phase node of grid entry g. In general, $\mathbf{R}_\theta(\boldsymbol{k}) \in \mathbb{C}^{d\times d}$ mixes channels at each retained mode. Our implementation uses the depthwise form $\mathbf{R}_\theta(\boldsymbol{k}) = \mathrm{diag}(\boldsymbol{r}_\theta(\boldsymbol{k}))$, which applies a learned complex multiplier to each channel. Retaining modes in conjugate pairs with $\mathbf{R}_\theta(-\boldsymbol{k}) = \overline{\mathbf{R}_\theta(\boldsymbol{k})}$ and using real multipliers on self-conjugate modes ensure real outputs for real inputs. The cutoff M determines the retained spatial frequencies. Because each Fourier mode spans the domain, this branch propagates broad spatial structure across the entire grid in one update. Lemma 7 shows that this finite-mode branch stays within $O(h_g)$ of its continuum counterpart as the grid is refined.

State-dependent interactions. The second path adapts spatial interactions to the current field features. Define $\boldsymbol{q}_g = \phi(Q_\theta(\boldsymbol{z}_g^{\mathrm{c}}))$, $\boldsymbol{k}_g = \phi(K_\theta(\boldsymbol{z}_g^{\mathrm{c}}))$, and $\boldsymbol{v}_g = V_\theta(\boldsymbol{z}_g^{\mathrm{c}})$, where $\phi(x) = \mathrm{ELU}(x) + 1$, $\boldsymbol{q}_g, \boldsymbol{k}_g \in \mathbb{R}^{d_a}$, and $\boldsymbol{v}_g \in \mathbb{R}^d$. Normalized linear attention (Katharopoulos et al., 2020) gives

$$\mathcal{A}_\theta(\mathbf{Z}^{\mathrm{c}})_g = O_\theta\left(\frac{\sum_{j=1}^{G}(\boldsymbol{q}_g^\top \boldsymbol{k}_j)\boldsymbol{v}_j}{\sum_{j=1}^{G}\boldsymbol{q}_g^\top \boldsymbol{k}_j + \epsilon_a}\right) = O_\theta\left(\frac{(\mathbf{V}^\top\mathbf{K})\boldsymbol{q}_g}{\boldsymbol{q}_g^\top\mathbf{K}^\top\mathbf{1}_G + \epsilon_a}\right), \tag{6}$$

where $\epsilon_a > 0$ is a numerical stabilizer and O_θ is an affine output map. The rows of $\mathbf{Q}, \mathbf{K} \in \mathbb{R}^{G\times d_a}$ and $\mathbf{V} \in \mathbb{R}^{G\times d}$ contain the queries, keys, and values. The second expression uses two aggregates shared by all queries: $\mathbf{V}^\top\mathbf{K} = \sum_{j=1}^{G} \boldsymbol{v}_j\boldsymbol{k}_j^\top \in \mathbb{R}^{d\times d_a}$ and $\mathbf{K}^\top\mathbf{1}_G = \sum_{j=1}^{G}\boldsymbol{k}_j \in \mathbb{R}^{d_a}$. Multiplying these aggregates by $\boldsymbol{q}_g$ recovers the weighted value sum and the normalization term in the first expression. Forming the aggregates costs $O(Gdd_a)$ and supplies all queries. The positive feature map produces positive interaction scores, whose scale is controlled by normalization. For bounded features and fixed neural maps, Lemma 4 establishes a Lipschitz bound independent of G.

The block output is

$$\mathbf{H} = \mathrm{LN}\big(\mathbf{Z} + \mathcal{P}_\theta(\mathbf{Z}^{\mathrm{c}}) + \mathcal{A}_\theta(\mathbf{Z}^{\mathrm{c}})\big) + \mathbf{1}_G\boldsymbol{b}_{\mathrm{res}}^\top, \tag{7}$$

where LN normalizes channels independently at each grid location and $\boldsymbol{b}_{\mathrm{res}} \in \mathbb{R}^d$ is a learned bias shared across the grid. Equivalently, $\boldsymbol{h}_g = \mathrm{LN}(\boldsymbol{z}_g + \mathcal{P}_\theta(\mathbf{Z}^{\mathrm{c}})_g + \mathcal{A}_\theta(\mathbf{Z}^{\mathrm{c}})_g) + \boldsymbol{b}_{\mathrm{res}}$. This update combines a direct feature-preserving path, global spectral propagation, and content-dependent spatial mixing. Their roles are complementary: spectral weights encode shared propagation patterns, and attention changes its affinities with the observed field. The unmodulated residual provides a reference for both updates. Lemma 5 combines the branchwise bounds, accounting for shared spectral conditioning, to establish stability of the full operator stack.

3.4 Coordinate Query Decoder

The decoder maps the evolved field to predictions at observed query sensors. Using the Gaussian kernel, define $b_{qg} = \kappa_{\sigma_{\mathrm{dec}}}(\boldsymbol{x}_q, \boldsymbol{\xi}_g)/\sum_{r=1}^{G}\kappa_{\sigma_{\mathrm{dec}}}(\boldsymbol{x}_q, \boldsymbol{\xi}_r)$ and gather $\tilde{\boldsymbol{h}}_q = \sum_{g=1}^{G} b_{qg}\boldsymbol{h}_g = \mathbf{H}^\top\boldsymbol{b}_q$, where $\boldsymbol{b}_q = [b_{q1}, \ldots, b_{qG}]^\top$ satisfies $b_{qg} \geq 0$ and $\mathbf{1}_G^\top\boldsymbol{b}_q = 1$. The learned decoder bandwidth $\sigma_{\mathrm{dec}} > 0$ controls the readout scale. Shared prediction models combine field and local-history contributions:

$$\hat{\boldsymbol{y}}_q = \underbrace{\mathbf{W}_{\mathrm{out}}\,\mathrm{LN}(\tilde{\boldsymbol{h}}_q) + \boldsymbol{b}_{\mathrm{out}}}_{\textit{field contribution } \boldsymbol{f}_q} + \underbrace{\mathbf{W}_{\mathrm{skip}}\boldsymbol{u}_q + \boldsymbol{b}_{\mathrm{skip}}}_{\textit{local-history contribution}}. \tag{8}$$

Here $\mathbf{W}_{\mathrm{out}} \in \mathbb{R}^{T_f\times d}$ and $\mathbf{W}_{\mathrm{skip}} \in \mathbb{R}^{T_f\times T_h}$. We initialize the local-history map with $\mathbf{W}_{\mathrm{skip}} = \mathbf{1}_{T_f}\boldsymbol{e}_{T_h}^\top$ and zero bias, where $\boldsymbol{e}_{T_h}$ is the last standard basis vector in $\mathbb{R}^{T_h}$. This initialization repeats the latest observation across all forecast horizons. This decomposition lets the local-history path express sensor-specific temporal persistence and the field path supply spatially informed corrections. Lemma 3 bounds the gather error of the normalized weights by $O(h_g)$.

3.5 Continual Update and Consistency

We train **STFO** from scratch in the first period. For each subsequent period $p > 1$, we initialize the model from $\theta^{(p-1)}$ and fine-tune the same parameter set on current training samples. New sensors enter through their histories and period-specific coordinates. The latent grid and all parameter

shapes, including those of the Fourier multipliers, remain fixed, allowing the learned operator to be reused and fine-tuned across periods. Algorithms 1 and 2 in Appendix E describe the per-period forward pass and continual training protocol, respectively. Appendix D provides the masked MAE objective and optimization settings.

The analysis follows this computational structure. Kernel-lift errors propagate through a Lipschitz post-network (Lemmas 1 and 2), a stable spectral descriptor (Theorem 3), a finite stack of operator blocks (Lemmas 4 and 5), and a normalized decoder (Lemma 3). With fixed bandwidth and a common limiting sampling measure, prediction discrepancy from the reference is bounded by the distance between sampling measures; with local bandwidth proportional to fill distance, it is $O(h_S)$. For a fixed spectral design and decoder bandwidth proportional to grid spacing, the grid error is $O(h_g)$, where the continuum comparison rests on endpoint-grid quadrature (Lemma 6) and on the grid stack (Lemma 7). Theorems 1 and 2 state the resulting bounds, Corollary 1 compares two layouts under a shared reference, and Appendix C.4 separates the effect of coordinate reassignment.

4 EXPERIMENTS

We investigate the following research questions:

- ***RQ1:*** How accurate is the proposed method across different datasets and horizons?
- ***RQ2:*** How much does each model component contribute to the overall forecasting performance?
- ***RQ3:*** What spatial and channel-wise structures characterize the learned field updates, and how do spectral conditioning and channel modulation relate to forecasting improvements?
- ***RQ4:*** How does the proposed method compare in terms of forecasting accuracy, training efficiency, and memory usage?
- ***RQ5:*** How well does the proposed method forecast at newly introduced sensors while retaining performance on existing sensors under different adaptation settings?
- ***RQ6:*** How sensitive is the proposed method to operator depth, grid resolution, and the number of retained modes? We report this sensitivity study in Appendix I.

4.1 EXPERIMENTAL SETUP

Datasets. We use PEMS-STREAM (Chen et al., 2021), CA-STREAM (Liu et al., 2023b), and AIR-STREAM (Chen & Liang, 2025a). PEMS-STREAM grows from 655 to 871 traffic sensors over seven yearly periods; CA-STREAM grows from 480 to $1{,}698$ sensors over four monthly periods; AIR-STREAM grows from $1{,}087$ to $1{,}202$ air-quality stations over four yearly periods.

Baselines. We group the baselines by modeling and adaptation paradigm: ❶ General spatio-temporal forecasting methods include GWNET (Wu et al., 2019), STID (Shao et al., 2022), and ITRANSFORMER (Liu et al., 2023c); ❷ Neural-operator baselines include DEEPONET (Lu et al., 2021) and FNO (Li et al., 2020); ❸ Continual-learning methods for evolving spatio-temporal systems include TRAFFICSTREAM (Chen et al., 2021), STKEC (Wang et al., 2023a), PECPM (Wang et al., 2023b), STRAP (Zhang et al., 2025a), EAC (Chen & Liang, 2025a), and EVOLVE (Yang et al., 2027); ❹ Online adaptation methods include DOL (Wang et al., 2025) and ST-TTC (Chen & Liang, 2025b), which update or calibrate predictions using streaming observations at deployment time. Appendix F describes each baseline and its role in the comparison.

Protocol and metrics. For subsequent periods, we initialize the model from the previous checkpoint and fine-tune on current training data. Sensor histories and coordinates reflect the current sensor set, while the latent grid and parameter shapes remain fixed. For each dataset, the Small, Medium, and Large configurations use different hidden widths. We report MAE, RMSE, and MAPE at horizons 3, 6, and 12, and their averages over 12 horizons. Main results are averaged over three runs with random seeds 42, 43, and 44. Appendices D and E provide experimental details and algorithms.

4.2 ALL-HORIZON FORECASTING RESULTS (RQ1)

Table 1 shows that **STFO**-Large achieves the lowest errors across all reported horizons and metrics on both traffic streams, reducing average MAE and RMSE relative to DOL by about 8% on PEMS-

Table 1: Forecasting results across datasets and horizons, with methods grouped by forecasting paradigm. Results are mean ± standard deviation over three runs. Lower is better; **bold** and underlined denote the best and second-best methods within each dataset, metric, and horizon.

Metric	Hor.	Traditional			Neural Operators		Continual Learning						Online Adaptation		STFO *(Ours)*		
		GWNet	STID	iTrans	DeepONet	FNO	TStream	STKEC	PECPM	STRAP	EAC	EVOLVE	DOL	ST-TTC	Small	Medium	Large
PEMS-Stream																	
MAE	3	19.64±0.12	24.34±0.13	17.63±0.76	12.25±0.02	12.63±0.02	14.23±0.09	14.29±0.12	14.26±0.13	13.09±0.01	13.86±0.16	13.30±0.09	11.40±0.03	12.94±0.11	11.26±0.02	10.94±0.02	**10.76**±0.02
	6	19.68±0.19	25.45±0.21	20.82±0.76	13.72±0.04	14.37±0.02	16.43±0.03	16.44±0.11	16.35±0.12	15.02±0.05	15.40±0.19	15.06±0.08	12.24±0.05	14.22±0.17	12.03±0.02	11.55±0.02	**11.27**±0.03
	12	20.63±0.09	29.42±0.38	28.33±0.86	16.90±0.03	18.16±0.02	21.76±0.07	21.66±0.11	21.46±0.19	19.78±0.00	18.90±0.28	19.50±0.09	13.86±0.10	17.36±0.25	13.64±0.03	12.81±0.06	**12.28**±0.06
	Avg.	19.87±0.10	26.07±0.23	21.60±0.79	13.96±0.02	14.68±0.02	16.95±0.03	16.96±0.09	16.86±0.12	15.47±0.01	15.67±0.20	15.50±0.08	12.33±0.06	14.47±0.16	12.11±0.02	11.61±0.03	**11.30**±0.03
RMSE	3	32.20±0.17	39.37±0.13	28.20±1.15	20.05±0.05	20.78±0.02	23.00±0.09	23.08±0.14	23.07±0.15	21.53±0.08	22.26±0.23	21.73±0.14	18.84±0.04	20.84±0.20	18.65±0.05	18.11±0.02	**17.83**±0.03
	6	32.34±0.32	40.86±0.19	33.80±1.13	22.50±0.04	23.97±0.04	26.87±0.04	26.93±0.15	26.76±0.20	24.86±0.16	24.99±0.28	24.72±0.14	20.37±0.09	22.96±0.30	19.99±0.06	19.18±0.04	**18.78**±0.04
	12	33.73±0.09	46.20±0.43	45.98±1.25	27.40±0.06	30.39±0.04	35.29±0.11	35.19±0.11	34.77±0.37	32.46±0.00	30.56±0.45	31.71±0.17	23.00±0.20	27.68±0.45	22.53±0.06	21.15±0.09	**20.40**±0.08
	Avg.	32.59±0.18	41.67±0.21	34.88±1.17	22.78±0.04	24.39±0.02	27.52±0.05	27.56±0.11	27.37±0.20	25.46±0.07	25.30±0.29	25.31±0.14	20.42±0.10	23.23±0.28	20.04±0.06	19.20±0.04	**18.76**±0.05
MAPE	3	27.47±0.69	37.79±2.23	32.46±3.04	19.67±0.22	20.35±0.04	18.34±0.67	18.54±0.61	18.19±0.66	17.05±0.02	18.35±0.31	17.77±0.28	15.41±0.05	20.69±0.53	15.07±0.02	14.71±0.07	**14.55**±0.14
	6	27.22±0.58	39.70±2.43	36.73±3.84	23.51±0.12	24.69±0.14	20.77±0.71	20.64±0.48	20.79±0.57	19.35±0.05	20.11±0.36	19.63±0.18	16.32±0.09	21.88±0.04	16.23±0.06	15.79±0.03	**15.29**±0.13
	12	29.38±1.18	47.94±2.91	54.31±4.66	33.32±0.11	35.63±0.09	27.88±0.26	27.05±0.62	28.33±0.52	25.46±0.52	24.30±0.57	24.81±0.13	18.30±0.14	26.61±0.93	18.79±0.31	17.90±0.22	**16.97**±0.10
	Avg.	27.79±0.76	41.09±2.49	39.63±3.81	24.61±0.13	25.90±0.03	21.66±0.54	21.50±0.52	21.73±0.45	20.05±0.18	20.42±0.41	20.29±0.16	16.47±0.09	22.38±0.21	16.45±0.11	15.89±0.09	**15.43**±0.09
CA-Stream																	
MAE	3	23.49±0.80	27.71±0.23	20.16±0.06	14.87±0.02	15.13±0.04	17.82±0.26	17.69±0.19	17.93±0.12	15.65±0.02	17.66±0.37	16.09±0.04	14.07±0.03	15.63±0.08	14.02±0.03	13.86±0.03	**13.61**±0.02
	6	23.31±0.69	28.93±0.26	24.37±0.06	16.60±0.04	17.05±0.03	20.38±0.17	20.41±0.04	20.33±0.09	18.00±0.03	19.68±0.54	18.30±0.03	15.22±0.04	17.22±0.08	15.12±0.02	14.89±0.07	**14.55**±0.03
	12	24.78±0.86	33.61±0.45	34.05±0.06	20.00±0.10	20.67±0.05	26.92±0.53	27.05±0.17	26.68±0.19	23.50±0.04	24.86±1.33	23.76±0.10	17.32±0.05	20.91±0.14	17.34±0.06	16.89±0.10	**16.29**±0.06
	Avg.	23.73±0.75	29.71±0.28	25.34±0.05	16.80±0.05	17.25±0.04	21.09±0.29	21.09±0.13	21.04±0.11	18.51±0.01	20.20±0.69	18.85±0.04	15.30±0.04	17.50±0.08	15.23±0.02	14.97±0.07	**14.58**±0.03
RMSE	3	35.87±0.98	41.53±0.31	31.58±0.09	24.99±0.04	25.29±0.05	28.01±0.22	28.02±0.19	28.00±0.16	26.37±0.11	27.46±0.46	26.96±0.03	24.21±0.02	25.55±0.07	24.00±0.05	23.78±0.06	**23.40**±0.03
	6	35.68±0.88	43.14±0.35	37.76±0.10	27.61±0.07	28.23±0.08	32.19±0.22	32.43±0.05	31.94±0.09	30.19±0.08	30.64±0.83	30.57±0.03	26.06±0.02	27.95±0.14	25.75±0.02	25.42±0.10	**24.91**±0.05
	12	37.57±1.11	49.18±0.58	51.24±0.10	32.34±0.12	33.50±0.07	41.59±0.64	42.08±0.21	41.14±0.30	38.57±0.09	37.77±1.94	38.67±0.09	29.06±0.04	33.04±0.24	28.92±0.08	28.34±0.14	**27.43**±0.10
	Avg.	36.20±0.96	44.12±0.37	38.94±0.09	27.80±0.09	28.47±0.06	33.01±0.35	33.24±0.13	32.77±0.17	30.87±0.06	31.18±0.99	31.25±0.04	26.09±0.03	28.23±0.13	25.83±0.04	25.47±0.09	**24.91**±0.04
MAPE	3	24.61±0.95	29.24±0.65	21.76±0.17	15.45±0.10	15.70±0.06	17.05±0.41	16.60±0.19	17.63±0.91	16.17±0.25	18.26±1.88	16.59±0.20	14.88±0.03	16.66±0.14	14.65±0.04	14.48±0.04	**14.34**±0.01
	6	24.44±0.80	30.66±0.78	26.76±0.22	17.05±0.13	17.45±0.11	19.22±0.30	18.98±0.17	19.74±0.92	18.46±0.36	19.45±1.16	18.73±0.20	15.83±0.06	18.63±0.29	15.79±0.08	15.42±0.03	**15.24**±0.13
	12	25.71±0.81	36.88±1.29	39.81±0.38	21.02±0.26	21.03±0.10	25.47±0.46	24.99±0.29	25.94±1.09	23.93±0.53	24.52±1.10	24.41±0.31	17.78±0.05	22.73±0.12	18.30±0.25	17.67±0.07	**17.07**±0.06
	Avg.	24.79±0.85	31.73±0.86	28.34±0.20	17.46±0.17	17.67±0.07	19.98±0.30	19.61±0.19	20.49±0.91	18.99±0.38	20.17±1.25	19.38±0.22	15.94±0.05	19.07±0.24	15.96±0.05	15.63±0.05	**15.30**±0.00
AIR-Stream																	
MAE	3	28.48±1.43	32.85±0.21	22.37±0.76	16.47±0.06	16.29±0.04	20.73±0.40	20.95±0.17	20.82±0.35	16.53±0.11	20.41±0.36	16.45±0.03	**15.70**±0.02	16.75±0.06	15.95±0.03	15.92±0.06	15.98±0.02
	6	29.79±0.89	33.15±0.22	26.22±0.48	22.18±0.16	21.66±0.14	25.64±0.34	25.54±0.08	25.54±0.19	22.87±0.01	25.20±0.29	22.66±0.03	21.33±0.05	22.68±0.04	21.24±0.03	**21.18**±0.08	21.26±0.01
	12	31.30±0.52	33.88±0.25	29.45±0.31	26.16±0.16	25.64±0.13	29.04±0.23	28.94±0.12	28.95±0.11	27.29±0.19	28.57±0.42	27.03±0.02	25.75±0.08	26.75±0.05	**25.27**±0.03	25.31±0.14	25.42±0.07
	Avg.	29.66±1.01	33.23±0.22	25.53±0.56	20.78±0.10	20.37±0.11	24.58±0.34	24.63±0.11	24.60±0.21	21.41±0.10	24.21±0.43	21.25±0.03	20.13±0.04	21.27±0.03	20.01±0.01	**19.99**±0.08	20.06±0.03
RMSE	3	44.38±2.04	51.24±0.28	34.98±1.18	27.93±0.09	27.32±0.12	32.80±0.57	33.13±0.28	33.07±0.52	28.05±0.14	32.19±0.57	28.08±0.05	27.08±0.05	27.74±0.09	26.98±0.04	**26.93**±0.11	26.96±0.05
	6	46.22±1.28	51.61±0.31	40.95±0.73	36.10±0.21	34.81±0.13	40.41±0.53	40.38±0.20	40.48±0.39	37.14±0.05	39.63±0.43	36.94±0.07	35.41±0.08	36.03±0.11	34.55±0.03	**34.36**±0.14	34.38±0.08
	12	48.34±0.85	52.55±0.39	45.70±0.55	41.67±0.39	40.58±0.18	45.54±0.47	45.53±0.27	45.63±0.27	43.64±0.34	44.65±0.63	43.21±0.07	41.57±0.13	41.73±0.17	40.30±0.08	40.30±0.17	**40.29**±0.11
	Avg.	46.01±1.46	51.72±0.33	39.67±0.91	33.98±0.18	32.98±0.14	38.58±0.53	38.70±0.26	38.76±0.41	34.97±0.11	37.83±0.60	34.80±0.07	33.45±0.09	33.94±0.09	32.71±0.05	**32.61**±0.11	32.63±0.07
MAPE	3	38.02±2.60	43.52±0.64	28.64±1.28	22.64±0.10	22.62±0.36	26.33±0.30	26.24±0.30	25.79±0.50	22.42±0.16	26.06±0.71	22.31±0.02	**21.40**±0.04	24.26±0.15	21.87±0.08	21.86±0.08	22.09±0.15
	6	39.98±1.70	44.12±0.57	34.91±0.68	30.88±0.35	30.54±0.58	33.33±0.21	33.10±0.28	32.97±0.18	31.31±0.10	32.88±0.64	31.51±0.04	**29.48**±0.11	33.55±0.26	29.52±0.15	29.80±0.22	29.91±0.39
	12	42.37±1.14	45.06±0.62	40.79±0.39	37.33±0.26	36.67±0.67	39.27±0.24	39.02±0.18	38.67±0.02	37.70±0.17	38.85±0.67	38.21±0.10	36.69±0.15	41.06±0.34	**35.82**±0.19	36.35±0.46	36.67±0.63
	Avg.	39.87±1.87	44.16±0.60	34.15±0.76	29.09±0.22	28.70±0.45	32.29±0.29	32.12±0.21	31.82±0.19	29.35±0.10	31.77±0.53	29.57±0.05	28.09±0.09	31.66±0.21	**27.90**±0.14	28.19±0.22	28.36±0.39

Stream and 5% on CA-Stream. Relative gains over DOL increase with the prediction horizon and are largest at horizon 12. On AIR-Stream, **STFO** also achieves the best average results across all three metrics. Scaling from Small to Large consistently improves performance on the two traffic streams, whereas improvements on AIR-Stream are not monotonic. Holm-corrected Wilcoxon signed-rank tests in Appendix G support **STFO**-Large's overall advantage over the baselines.

4.3 Ablation Study (RQ2)

Figure 3 compares **STFO**-Large with four ablated variants on PEMS-Stream and AIR-Stream: w/o SRE removes spectral conditioning, w/o DFO-S removes the spectral path, w/o DFO-A removes the attention path, and w/o CFE-C removes coordinate-feature injection. The full model has the lowest MAE and MAPE in both panels. Removing coordinate features produces the largest MAE increase on both streams and the largest MAPE increase on PEMS-Stream, highlighting the role of explicit location information in the lifted field. On AIR-Stream, removing the spectral path causes the largest MAPE increase but the smallest MAE increase among the ablations. These differing rankings suggest that component contributions depend on the dataset and evaluation metric. The performance degradation after removing either DFO path further suggests that both spectral propagation and state-dependent interactions contribute to forecasting accuracy.

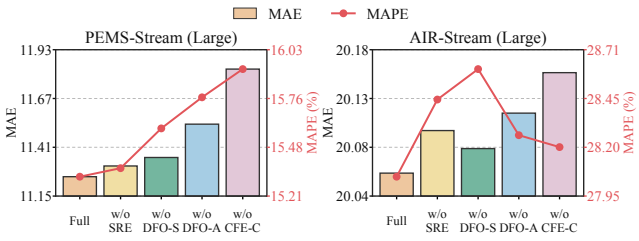

Figure 3: Component ablation of **STFO**-Large on PEMS-Stream and AIR-Stream. Results are from single-seed runs.

4.4 Mechanism Diagnostics (RQ3)

❶ ***At the spatial level,*** forecasting gains and latent-field updates exhibit spatially structured patterns that evolve differently across datasets. Figures 4 and 5 show error reductions relative to the local-

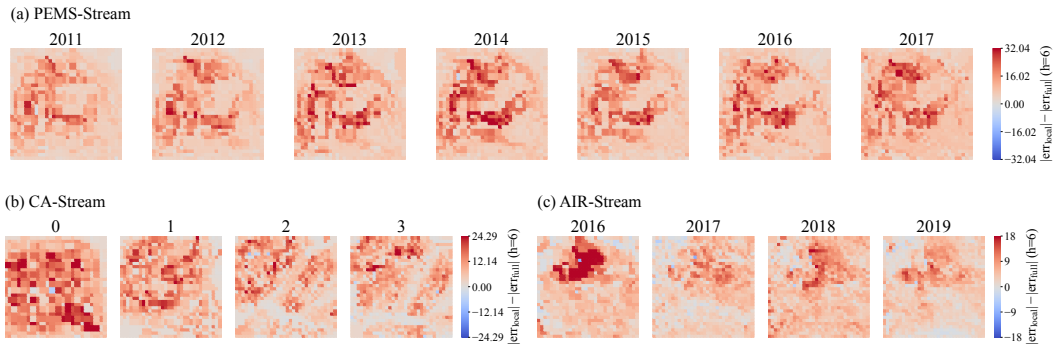

Figure 4: Sensor-wise error reduction $|\mathrm{err}_{\mathrm{local}}| - |\mathrm{err}_{\mathrm{full}}|$ at horizon 6, shown on the assignment grid for all periods of the three streams. Here, $\mathrm{err}_{\mathrm{local}}$ denotes the error of the local-history-only prediction. Positive values indicate improvements from adding the field contribution.

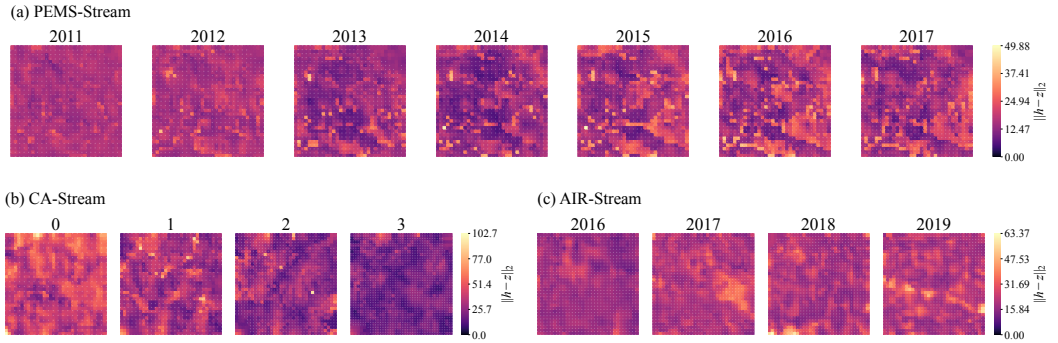

Figure 5: DFO update magnitudes on the latent grid across three streams and all periods. Each location g shows $\|h_g - z_g\|_2$, where z_g and h_g denote the input and output latent features, respectively.

history-only prediction at horizon 6 and latent-field update magnitudes, respectively. On PEMS-Stream, forecasting gains remain broadly distributed across the coordinate domain, while field updates exhibit more pronounced local peaks in later periods. CA-Stream shows more visible changes in the spatial distribution of forecasting gains, while field-update magnitudes tend to decrease across periods. On AIR-Stream, both forecasting gains and field updates display spatial structure, with their locations and intensities varying across years. These observations suggest that the shared field operator accommodates both persistent spatial structure and changes across periods, with the field contribution improving predictions at many sensor locations.

❷ ***At the channel level,*** spectral conditioning produces modulation patterns that differ across channels and periods. Figure 6 shows positive and negative correlations between SRE descriptor components and the FiLM coefficients $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$. These patterns indicate that latent channels respond differently to the spectral regime. The analysis in Figure 9 of Appendix J further indicates that the relationship between channel modulation and forecasting gains varies across datasets and periods.

4.5 Efficiency Analysis (RQ4)

Figure 7 compares forecasting accuracy, training time per epoch, and GPU memory usage on PEMS-Stream. All three **STFO** configurations achieve lower MAE than the plotted baselines, offering favorable accuracy–cost trade-offs across model sizes. **STFO**-Small achieves strong forecasting performance with 2.46 seconds per epoch and 0.38 GiB of GPU memory. Increasing model capacity further improves accuracy, with **STFO**-Large achieving the lowest MAE at 3.22 seconds per epoch and 1.02 GiB

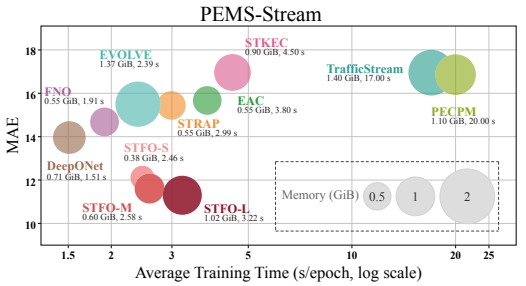

Figure 7: Forecasting accuracy, training time per epoch, and GPU memory usage on PEMS-Stream.

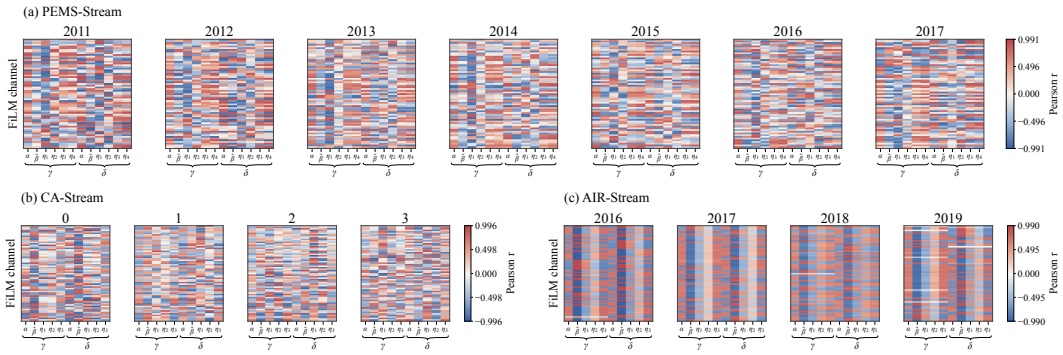

Figure 6: SRE-to-FiLM channel correlations across three streams. The heatmaps show correlations between SRE descriptor components and channel-wise FiLM parameters γ and δ.

Table 2: Period-wise MAE on existing and newly introduced sensors of AIR-Stream. Results are mean $\pm$ standard deviation over three runs. Lower is better; **bold** and underlined denote the best and second-best methods within each period, sensor group, adaptation budget, and horizon.

Budget	Sensor group	FNO				EVOLVE				STFO-Large			
		$h=3$	$h=6$	$h=12$	*Avg.*	$h=3$	$h=6$	$h=12$	*Avg.*	$h=3$	$h=6$	$h=12$	*Avg.*
2017: 1,087 existing sensors; 67 newly introduced sensors													
Zero-shot	Existing	17.57±0.26	23.15±0.19	27.09±0.14	21.71±0.15	17.37±0.05	23.92±0.03	28.38±0.06	22.39±0.04	**16.93**±0.14	**22.54**±0.26	**26.91**±0.39	**21.26**±0.26
	Newly introduced	18.12±0.32	24.11±0.15	29.13±0.34	22.80±0.24	18.12±0.07	25.53±0.05	31.46±0.08	24.10±0.06	**17.36**±0.18	**23.41**±0.34	**28.82**±0.45	**22.25**±0.33
Full	Existing	16.93±0.11	22.53±0.17	26.56±0.21	21.15±0.14	17.32±0.21	23.86±0.12	28.36±0.12	22.36±0.17	**16.61**±0.05	**22.06**±0.04	**26.24**±0.03	**20.80**±0.02
	Newly introduced	17.32±0.14	23.40±0.21	28.61±0.30	22.17±0.19	17.96±0.19	25.36±0.12	31.44±0.14	24.01±0.18	**16.96**±0.07	**22.79**±0.08	**28.07**±0.09	**21.68**±0.07
2018: 1,154 existing sensors; 39 newly introduced sensors													
Zero-shot	Existing	14.56±0.09	19.56±0.21	23.10±0.19	18.42±0.17	15.16±0.39	20.41±0.19	23.99±0.07	19.22±0.25	**14.41**±0.06	**19.26**±0.07	**22.90**±0.09	**18.17**±0.05
	Newly introduced	15.12±0.14	20.78±0.35	24.61±0.30	19.54±0.32	15.61±0.43	21.06±0.22	24.86±0.10	19.87±0.28	**14.67**±0.08	**19.71**±0.14	**23.52**±0.07	**18.60**±0.06
Full	Existing	14.39±0.07	19.23±0.15	22.81±0.08	18.09±0.06	14.78±0.04	20.47±0.13	24.35±0.19	19.18±0.10	**14.17**±0.00	**19.03**±0.01	**22.59**±0.05	**17.87**±0.02
	Newly introduced	14.95±0.06	20.45±0.06	24.48±0.32	19.22±0.10	15.16±0.03	21.09±0.13	25.25±0.23	19.79±0.11	**14.48**±0.03	**19.56**±0.06	**23.32**±0.05	**18.39**±0.03
2019: 1,193 existing sensors; 9 newly introduced sensors													
Zero-shot	Existing	15.38±0.18	19.83±0.30	23.18±0.12	18.85±0.12	15.36±0.49	20.71±0.29	24.38±0.23	19.53±0.35	**15.11**±0.06	**19.57**±0.13	**23.09**±0.11	**18.66**±0.08
	Newly introduced	21.14±0.46	**26.97**±0.58	**30.46**±0.48	**25.29**±0.40	21.22±0.68	27.98±0.31	31.51±0.19	26.12±0.34	**21.07**±0.08	27.08±0.29	31.24±0.53	25.61±0.29
Full	Existing	14.13±0.13	18.98±0.26	22.58±0.12	17.85±0.16	14.55±0.08	20.17±0.03	24.06±0.04	18.89±0.06	**14.08**±0.09	**18.78**±0.10	**22.37**±0.06	**17.69**±0.05
	Newly introduced	20.18±0.17	**26.31**±0.28	**29.85**±0.10	**24.46**±0.14	20.47±0.17	27.61±0.15	31.62±0.29	25.65±0.21	**20.10**±0.03	26.41±0.13	30.87±0.32	24.80±0.14

of memory. The larger configurations therefore improve accuracy while retaining competitive computational efficiency. Across these configurations, capacity increases through hidden width without enlarging the latent grid, while linear attention avoids constructing and storing a full grid-to-grid attention matrix. Combining a fixed latent grid with linear attention supports increased model capacity through wider hidden representations while keeping spatial computation efficient.

4.6 New-Sensor Adaptation (RQ5)

Table 2 compares FNO, EVOLVE, and **STFO**-Large on AIR-Stream under zero-shot and full adaptation. Without current-period fine-tuning, **STFO**-Large outperforms the baselines in most new-sensor comparisons, supporting its ability to transfer forecasting knowledge to newly introduced sensors. Full adaptation further reduces MAE for both new and existing sensors in every expansion period. Under both settings, **STFO**-Large achieves the lowest MAE in all existing-sensor comparisons and most new-sensor comparisons. These results suggest that knowledge transferred from previous periods provides a useful starting point for new sensors, while current-period adaptation further improves forecasting performance across the expanded network. Appendix H provides additional results across datasets and metrics.

5 Conclusion

In this work, we propose **STFO**, combining coordinate-based lifting and querying with spectrally conditioned Fourier and attention updates in a shared latent field. This formulation accommodates sensor expansion without changing parameter shapes, with prediction consistency established under sampling and grid-refinement assumptions. Experiments demonstrate leading accuracy on both traffic streams and competitive performance on AIR-Stream, supporting the shared latent field as an effective representation for continual forecasting under evolving sensor layouts and dynamics.

AI USE STATEMENT

Generative AI tools assisted with language revision.

REFERENCES

Ahmed Abdeljawad and Thomas Dittrich. Approximation rates in frechet metrics: Barron spaces, paley-wiener spaces, and fourier multipliers. *Neural Computation*, 2026. doi: 10.1162/NECO.a.1481. URL https://pubmed.ncbi.nlm.nih.gov/41370746/.

Sio-Iong Ao and Haytham M. Fayek. Continual deep learning for time series modeling. *Sensors*, 23 (16):7167, 2023. doi: 10.3390/s23167167. URL https://www.mdpi.com/1424-8220/23/16/7167/pdf?version=1692015651.

Pengfei Bie, Ning Song, Nuoqing Zhang, Jie Nie, Min Ye, Xinyue Liang, and Qi Wen. Space-frequency cross-attention node feature optimization graph neural operator for partial differential equations. *IEEE Transactions on Neural Networks and Learning Systems*, 2025. doi: 10.1109/TNNLS.2025.3643632. URL https://pubmed.ncbi.nlm.nih.gov/41447496/.

Jiawen Chen, Qi Shao, Duxin Chen, and Wenwu Yu. Decoupling spatio-temporal prediction: When lightweight large models meet adaptive hypergraphs. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2*, pp. 167–178, 2025. doi: 10.1145/3711896.3736904. URL https://arxiv.org/pdf/2505.19620.

Wei Chen and Yuxuan Liang. Expand and compress: Exploring tuning principles for continual spatio-temporal graph forecasting. In *The Thirteenth International Conference on Learning Representations*, 2025a.

Wei Chen and Yuxuan Liang. Learning with calibration: Exploring test-time computing of spatio-temporal forecasting. In *Advances in Neural Information Processing Systems*, volume 38, 2025b. doi: 10.52202/085713-5210.

Xu Chen, Junshan Wang, and Kunqing Xie. TrafficStream: A streaming traffic flow forecasting framework based on graph neural networks and continual learning. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*, pp. 3620–3626, 2021. doi: 10.24963/ijcai.2021/498. URL https://www.ijcai.org/proceedings/2021/0498.pdf.

Salvatore Cuomo, Vincenzo Schiano Di Cola, Fabio Giampaolo, Gianluigi Rozza, Maziar Raissi, and Francesco Piccialli. Scientific machine learning through physics-informed neural networks: Where we are and what's next. *Journal of Scientific Computing*, 92(3), 2022. doi: 10.1007/s10915-022-01939-z. URL https://link.springer.com/content/pdf/10.1007/s10915-022-01939-z.pdf.

Waleed Diab and Mohammed Al Kobaisi. Temporal neural operator for modeling time-dependent physical phenomena. *Scientific Reports*, 2025. doi: 10.1038/s41598-025-16922-5. URL https://pubmed.ncbi.nlm.nih.gov/40998917/.

Mariam Labib Francies, Abeer T. Khalil, Hanan M. Amer, and Mohamed Maher Ata. A framework for continual learning in real-time traffic forecasting utilizing spatial-temporal graph convolutional recurrent networks. *Complex & Intelligent Systems*, 11(10), 2025. doi: 10.1007/s40747-025-02049-7. URL https://link.springer.com/content/pdf/10.1007/s40747-025-02049-7.pdf.

Nicola Rares Franco, Andrea Manzoni, and Paolo Zunino. Mesh-informed neural networks for operator learning in finite element spaces. *Journal of Scientific Computing*, 97(2), 2023. doi: 10.1007/s10915-023-02331-1. URL https://link.springer.com/content/pdf/10.1007/s10915-023-02331-1.pdf.

Lukas Galke, Iacopo Vagliano, Benedikt Franke, Tobias Zielke, Marcel Hoffmann, and Ansgar Scherp. Lifelong learning on evolving graphs under the constraints of imbalanced classes and new classes. *Neural Networks*, 164:156–176, 2023. doi: 10.1016/j.neunet.2023.04.022. URL https://pure.amsterdamumc.nl/files/168252763/Lifelong-learning-o

n-evolving-graphs-under-the-constraints-of-imbalanced-classes-and-new-classes.pdf.

Shengnan Guo, Youfang Lin, Ning Feng, Chao Song, and Huaiyu Wan. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01):922–929, 2019. doi: 10.1609/aaai.v33i01.3301922. URL https://aaai.org/ojs/index.php/AAAI/article/download/3881/3759.

George Em Karniadakis, Ioannis G. Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021. doi: 10.1038/s42254-021-00314-5. URL https://www.osti.gov/servlets/purl/2282016.

Angelos Katharopoulos, Apoorv Vyas, Nikolaos Pappas, and François Fleuret. Transformers are RNNs: Fast autoregressive transformers with linear attention. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 5156–5165. PMLR, 2020. URL https://proceedings.mlr.press/v119/katharopoulos20a.html.

Fuxian Li, Jie Feng, Huan Yan, Guangyin Jin, Fan Yang, Funing Sun, Depeng Jin, and Yong Li. Dynamic graph convolutional recurrent network for traffic prediction: Benchmark and solution. *ACM Transactions on Knowledge Discovery from Data*, 17(1):1–21, 2022. doi: 10.1145/3532611. URL https://dl.acm.org/doi/pdf/10.1145/3532611.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew M. Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895*, 2020. doi: 10.48550/arxiv.2010.08895. URL https://arxiv.org/abs/2010.08895.

Zongyi Li, Nikola Kovachki, Chris Choy, Boyi Li, Jean Kossaifi, Shourya Otta, Mohammad Amin Nabian, Maximilian Stadler, Christian Hundt, Kamyar Azizzadenesheli, and Animashree Anandkumar. Geometry-informed neural operator for large-scale 3d pdes. In *Advances in Neural Information Processing Systems*, volume 36, pp. 35836–35854. Curran Associates, Inc., 2023. doi: 10.52202/075280-1556. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/70518ea42831f02afc3a2828993935ad-Paper-Conference.pdf.

Aoyu Liu and Yaying Zhang. A general spatio-temporal backbone with scalable contextual pattern bank for urban continual forecasting. In *International Conference on Learning Representations*, 2026. URL https://github.com/Aoyu-Liu/STBP.

Hangchen Liu, Zheng Dong, Renhe Jiang, Jiewen Deng, Jinliang Deng, Quanjun Chen, and Xuan Song. STAEformer: Spatio-temporal adaptive embedding makes vanilla transformer sota for traffic forecasting. *arXiv preprint arXiv:2308.10425*, 2023a. doi: 10.48550/arxiv.2308.10425. URL https://arxiv.org/abs/2308.10425.

Xu Liu, Yutong Xia, Yuxuan Liang, Junfeng Hu, Yiwei Wang, Lei Bai, Chao Huang, Zhenguang Liu, Bryan Hooi, and Roger Zimmermann. LargeST: A benchmark dataset for large-scale traffic forecasting. In *Advances in Neural Information Processing Systems*, volume 36, 2023b. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/ee57cd73a76bd927ffca3dda1dc3b9d4-Abstract-Datasets_and_Benchmarks.html.

Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. iTransformer: Inverted transformers are effective for time series forecasting. *arXiv preprint arXiv:2310.06625*, 2023c. doi: 10.48550/arxiv.2310.06625. URL https://arxiv.org/abs/2310.06625.

Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via deeponet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 3(3):218–229, 2021. doi: 10.1038/s42256-021-00302-5. URL https://arxiv.org/pdf/1910.03193.

Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. NeRF: representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021. doi: 10.1145/3503250. URL https://dl.acm.org/doi/pdf/10.1145/3503250.

Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics*, 41(4):1–15, 2022. doi: 10.1145/3528223.3530127. URL https://dl.acm.org/doi/pdf/10.1145/3528223.3530127.

Ethan Perez, Florian Strub, Harm de Vries, Vincent Dumoulin, and Aaron Courville. FiLM: Visual reasoning with a general conditioning layer. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1), 2018. doi: 10.1609/aaai.v32i1.11671. URL https://ojs.aaai.org/index.php/AAAI/article/view/11671.

Xihao Piao, Zheng Chen, Taichi Murayama, Yasuko Matsubara, and Yasushi Sakurai. Fredformer: Frequency debiased transformer for time series forecasting. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 2400–2410, 2024. doi: 10.1145/3637528.3671928. URL https://arxiv.org/pdf/2406.09009.

Zezhi Shao, Zhao Zhang, Fei Wang, Wei Wei, and Yongjun Xu. Spatial-temporal identity: A simple yet effective baseline for multivariate time series forecasting. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, pp. 4454–4458. Association for Computing Machinery, 2022. doi: 10.1145/3511808.3557702.

Jin Young Shin, Jae Yong Lee, and Hyung Ju Hwang. Pseudo-differential neural operator: Generalized fourier neural operator for learning solution operators of partial differential equations. *arXiv preprint arXiv:2201.11967*, 2022. doi: 10.48550/arxiv.2201.11967. URL https://arxiv.org/abs/2201.11967.

Chao Song, Youfang Lin, Shengnan Guo, and Huaiyu Wan. Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(01):914–921, 2020. doi: 10.1609/aaai.v34i01.5438. URL https://ojs.aaai.org/index.php/AAAI/article/download/5438/5294.

Matthew Tancik, Pratul Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. In *Advances in Neural Information Processing Systems*, volume 33, pp. 7537–7547. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/55053683268957697aa39fba6f231c68-Paper.pdf.

Bin-Wu Wang, Pengkun Wang, Yudong Zhang, Xu Wang, Zhengyang Zhou, Lei Bai, and Yang Wang. Towards dynamic spatial-temporal graph learning: A decoupled perspective. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(8):9089–9097, 2024. doi: 10.1609/aaai.v38i8.28759. URL https://ojs.aaai.org/index.php/AAAI/article/download/28759/29459.

Binwu Wang, Yudong Zhang, Jiahao Shi, Pengkun Wang, Xu Wang, Lei Bai, and Yang Wang. Knowledge expansion and consolidation for continual traffic prediction with expanding graphs. *IEEE Transactions on Intelligent Transportation Systems*, 24(7):7190–7201, 2023a.

Binwu Wang, Yudong Zhang, Xu Wang, Pengkun Wang, Zhengyang Zhou, Lei Bai, and Yang Wang. Pattern expansion and consolidation on evolving graphs for continual traffic prediction. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 2223–2232, 2023b.

Chengxin Wang, Gary Tan, Swagato Barman Roy, and Beng Chin Ooi. Distribution-aware online learning for urban spatiotemporal forecasting on streaming data. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, pp. 3344–3352, 2025. doi: 10.24963/ijcai.2025/372.

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. URL https://www.ijcai.org/proceedings/2019/0264.pdf.

Zonghan Wu, Shirui Pan, Guodong Long, Jing Jiang, Xiaojun Chang, and Chengqi Zhang. Connecting the dots: Multivariate time series forecasting with graph neural networks. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 753–763, 2020. doi: 10.1145/3394486.3403118. URL https://research-repository.griffith.edu.au/bitstreams/45e89825-c592-4b81-b421-f7327e69331e/download.

Shiyu Yang, Hengyu Guo, Chuxi Fan, Qunyong Wu, Mengmeng Li, and Zhiyuan Zhao. Evolve: A continual learning framework for evolving traffic networks. *Information Processing & Management*, 64(1):104974, 2027. doi: 10.1016/j.ipm.2026.104974.

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. URL https://www.ijcai.org/proceedings/2018/0505.pdf.

Qiao Yuan, Sheng-Uei Guan, Pin Ni, Tianlun Luo, Prudence Wong, Victor Chang, and Ka Lok Man. Continual graph learning: A survey. *Pattern Recognition*, 179:113600, 2026. doi: 10.1016/j.patcog.2026.113600. URL https://arxiv.org/pdf/2301.12230.

Haoyu Zhang, Wentao Zhang, Hao Miao, Xinke Jiang, Yuchen Fang, and Yifan Zhang. STRAP: Spatio-temporal pattern retrieval for out-of-distribution generalization. In *Advances in Neural Information Processing Systems*, volume 38, pp. 118006–118041, 2025a.

Haoyu Zhang, Xinke Jiang, Wentao Zhang, Yuchen Fang, Hao Miao, Yifan Zhang, and Heqing Huang. Pattern in motion: Retrieval-augmented learning for dynamic spatio-temporal graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–18, 2026. doi: 10.1109/TPAMI.2026.3728522.

Qianru Zhang, Xinyi Gao, Haixin Wang, Siu-Ming Yiu, and Hongzhi Yin. Efficient traffic prediction through spatio-temporal distillation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(1):1093–1101, 2025b. doi: 10.1609/aaai.v39i1.32096. URL https://ojs.aaai.org/index.php/AAAI/article/download/32096/34251.

Chuanpan Zheng, Xiaoliang Fan, Cheng Wang, and Jianzhong Qi. GMAN: A graph multi-attention network for traffic prediction. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(01):1234–1241, 2020. doi: 10.1609/aaai.v34i01.5477. URL https://ojs.aaai.org/index.php/AAAI/article/download/5477/5333.

CONTENTS OF THE APPENDIX

A Related Work **14**

B Notation and Symbols **15**

C Full Proofs **17**

C.1 Operator Conventions and Assumptions 17

C.2 Normalized Kernel Lifting and Gathering 18

C.3 Stability of the Trained Field Operator 20

C.4 Sensor-Discretization Consistency 21

C.5 Fixed-Design Grid-to-Continuum Consistency 22

C.6 Perturbation Stability of the Spectral State 24

D Experimental Setup Details **25**

E Algorithms **25**

F Baseline Descriptions **25**

G Statistical Significance Analysis **27**

H Full New-Sensor Adaptation Results **27**

I Hyperparameter Sensitivity **32**

J Additional Visual Diagnostics **33**

A RELATED WORK

Spatio-temporal forecasting models. Graph-based models have played a central role in spatio-temporal forecasting. Representative architectures combine graph convolution with gated temporal convolution (Yu et al., 2018), integrate dynamic graph construction with recurrent modeling (Li et al., 2022), learn adaptive adjacency matrices (Wu et al., 2019), construct synchronous space–time graphs (Song et al., 2020), learn directed graphs with mix-hop propagation (Wu et al., 2020), and apply spatial and temporal attention (Guo et al., 2019; Zheng et al., 2020). Other approaches show that competitive forecasting does not always require explicit graph construction: a vanilla transformer with adaptive spatio-temporal embeddings can achieve strong performance (Liu et al., 2023a), while inverted-attention and frequency-debiased transformers further advance sequence-based forecasting (Liu et al., 2023c; Piao et al., 2024). Recent work also pursues efficiency through spatio-temporal distillation (Zhang et al., 2025b) and lightweight language models coupled with adaptive hypergraphs (Chen et al., 2025). Despite differences in architecture, these methods generally organize spatial computation around observed sensors or variables, so changes in the observation set affect the representation used for forecasting.

Continual spatio-temporal forecasting. Continual forecasting methods extend spatio-temporal backbones to retain useful knowledge as data streams evolve and new sensors are introduced. Chen et al. (2021) combine experience replay with parameter smoothing for expanding road networks; subsequent work introduces knowledge expansion and consolidation (Wang et al., 2023a), pattern expansion and consolidation (Wang et al., 2023b), and continual prompt pools (Chen & Liang, 2025a). Meanwhile, Wang et al. (2024) decouple trend and seasonal patterns and update them at

different frequencies to adapt to evolving graphs. Continual learning has also been studied on evolving graphs (Yuan et al., 2026; Galke et al., 2023), while Ao & Fayek (2023) review continual deep learning for time-series modeling more broadly. Retrieval-augmented approaches store and retrieve historical patterns to support forecasting under distribution shifts in streaming spatio-temporal data (Zhang et al., 2026). These studies address adaptation to changing data and graph structure; our work focuses on the dependence of forecasting representations on sensor discretization.

Neural operators and implicit representations. Operator learning approximates mappings between functions. The Fourier neural operator parameterizes translation-invariant integral kernels in the spectral domain and supports resolution transfer for parametric PDEs (Li et al., 2020), while DeepONet represents operators through a branch–trunk architecture with universal approximation guarantees (Lu et al., 2021). Later work extends spectral operators (Shin et al., 2022), models time-dependent dynamics (Diab & Kobaisi, 2025), combines spectral and attention mechanisms (Bie et al., 2025), develops mesh-informed architectures for operator learning in finite-element spaces (Franco et al., 2023), and advances operator approximation theory (Abdeljawad & Dittrich, 2026). Physics-informed learning provides complementary tools for incorporating physical constraints (Karniadakis et al., 2021; Cuomo et al., 2022). In parallel, implicit neural representations encode continuous signals as functions of coordinates, as exemplified by neural radiance fields and multiresolution hash encodings (Mildenhall et al., 2021; Müller et al., 2022). GINO (Li et al., 2023) uses graph neural operators to map between irregular meshes and a regular latent grid, with a Fourier neural operator acting on the grid. Our work focuses on continual forecasting from expanding real-world sensor sets under distribution drift. We combine normalized Gaussian lifting and readout with spectral-regime conditioning and establish sensor-discretization consistency under explicit sampling assumptions.

B Notation and Symbols

Table 3 summarizes the symbols used in the main text and proofs. Scalars and scalar fields are italic (T_h, u_t), vectors are bold italic ($\boldsymbol{u}_i$, $\boldsymbol{q}_g$), matrices and gridded fields are bold upright ($\mathbf{Z}$, $\mathbf{A}$), and number spaces are blackboard bold ($\mathbb{R}$). Calligraphic letters denote selected sets and operators ($\mathcal{D}$, $\mathcal{S}$, $\mathcal{G}_\theta$, $\mathcal{P}_\theta$). In the proofs, dark blue marks continuum or reference objects, while dark orange marks discrete or empirical objects, as described at the start of Appendix C.

Table 3: Symbols used throughout the paper.

Symbol	Meaning
Problem and data	
P	Number of periods in the continual forecasting stream.
$\mathcal{V}_p$, N_p	Sensor set in period p and its size, $N_p = \lvert\mathcal{V}_p\rvert$; $\mathcal{V}_p \subseteq \mathcal{V}_{p+1}$ for $p < P$.
$\boldsymbol{s}_i$	Static descriptor vector of sensor i, containing geographic location and available attributes.
χ_p, $\boldsymbol{x}_i^{(p)}$	Period-specific coordinate map, the coordinate assigned to sensor i.
$\Omega = [0,1]^2$	Normalized spatial domain.
T_h, T_f	Input history length and forecast horizon.
u_t, $y_{i,t}$, $\varepsilon_{i,t}$	Spatial field, its noisy observation at sensor i, and observation noise.
$\boldsymbol{u}_i$	Observation history vector of sensor i.
$\mathcal{D}$	Input set of coordinate–history pairs for one period.
$\mathcal{G}_\theta$, $\hat{\boldsymbol{y}}_q$	Forecasting operator and its forecast at query sensor q.
$\mathcal{S}$, $h_{\mathcal{S}}$	Sensor layout and its fill distance.
$\nu_{\mathcal{S}}$, μ	Empirical sensor measure and the common limiting measure in fixed-bandwidth sampling comparisons.
$W_1(\cdot,\cdot)$	Wasserstein-1 distance between sampling measures.
Grid, field, and encoder	
n_g, $G = n_g^2$, $h_g = 1/(n_g - 1)$	Grid points per axis (denoted n in the proofs), total point count, and spacing.
Ξ_{h_g}, $\boldsymbol{\xi}_g$	Latent Cartesian grid and its node indexed by g.

continued on the next page

Table 3, continued from the previous page.

Symbol	Meaning
$\boldsymbol{t}_g$, $\boldsymbol{t}_{\boldsymbol{a}}$	Fourier phase node under scalar index g or grid multi-index $\boldsymbol{a}$.
d	Hidden width, i.e., the channel dimension of the latent field.
d_a	Dimension of attention queries and keys.
d_ρ	Output dimension of the Fourier coordinate encoding.
τ_θ, $\boldsymbol{e}_i$	Shared temporal encoder and the history code of sensor i.
κ_σ, σ_{enc}	Gaussian kernel and learned encoder bandwidth.
a_{gi}, $\mathbf{A}$	Normalized lift weights and their $G \times N$ matrix.
$\mathbf{E}$, $\mathbf{R}$	Sensor-code matrix and the lifted field $\mathbf{R} = \mathbf{A}\mathbf{E}$.
ρ, ψ_θ	Fixed Fourier coordinate encoding and shared post-encoding map.
$\mathbf{Z}$, $\boldsymbol{z}_g$	Latent grid field and the feature vector at node g; row g of $\mathbf{Z}$ is $\boldsymbol{z}_g^\top$.
Spectral regime descriptor	
$\bar{z}_g$, $\bar{z}$, $\widehat{\bar{z}}$	Channel mean at node g, the scalar grid, and its forward-normalized DFT.
$\mathcal{J}$, $\boldsymbol{\omega}_m$	Selected nonzero frequency bins and their wavevectors; spectral proofs use $\boldsymbol{k}_j$ with bin index j.
E_m, ω_m	Stabilized bin power and radial frequency; denoted E_j and k_j in spectral proofs.
ϵ_s, ϵ_f	Power/denominator and frequency stabilizers; written $\varepsilon_s, \varepsilon_f$ in the proofs.
λ_m, μ_m	Log radial frequency and log spectral power of bin m.
α, β, $\tilde{\beta}$	Normalized spectral centroid, regularized negative log–log slope, and clipped slope $\mathrm{clip}(\beta, -10, 10)$.
η_b, $\mathcal{B}_b$, B	Band-energy ratio, logarithmically spaced frequency band, and number of bands.
$\boldsymbol{c}$	Spectral regime descriptor $[\alpha, \tilde{\beta}, \eta_1, \ldots, \eta_B]^\top$.
$\mathcal{K}_M$, M	Retained low-frequency mode set and its cutoff.
$\mathbf{R}_\theta(\boldsymbol{k})$	Complex Fourier multiplier at each mode; diagonal in the depthwise implementation.
Field operator and decoder	
$\boldsymbol{\gamma}$, $\boldsymbol{\delta}$	FiLM gain adjustment and offset from $\boldsymbol{c}$; the multiplicative gain is $\mathbf{1}_d + \boldsymbol{\gamma}$.
$\mathbf{Z}^{\mathrm{c}}$	Channel-modulated latent field.
$\mathcal{P}_\theta$, $\mathcal{A}_\theta$	Spectral propagation path and normalized attention path.
ϕ, ϵ_a	Positive feature map $\mathrm{ELU}(\cdot) + 1$ and denominator stabilizer; the latter is written ε_a in the proofs.
$\mathbf{Q}$, $\mathbf{K}$, $\mathbf{V}$	Query, key, and value matrices, with rows $\boldsymbol{q}_g^\top$, $\boldsymbol{k}_g^\top$, and $\boldsymbol{v}_g^\top$, respectively.
O_θ	Affine output map of the attention path.
LN, $\boldsymbol{b}_{\mathrm{res}}$	Rowwise LayerNorm and the shared output bias of an operator block.
$\mathbf{H}$, $\boldsymbol{h}_g$	Block output field and the feature vector at node g; row g of $\mathbf{H}$ is $\boldsymbol{h}_g^\top$.
L	Number of stacked operator blocks.
b_{qg}, $\boldsymbol{b}_q$, σ_{dec}	Normalized decoder gather weights, their column vector, and the learned decoder bandwidth.
$\tilde{\boldsymbol{h}}_q$	Gathered field feature at query sensor q.
$\mathbf{W}_{\mathrm{out}}$, $\mathbf{W}_{\mathrm{skip}}$	Weight matrices of the field readout and local-history prediction.
$\boldsymbol{e}_{T_h}$	Last standard basis vector in $\mathbb{R}^{T_h}$, used to initialize $\mathbf{W}_{\mathrm{skip}}$.
$\boldsymbol{f}_q$	Field contribution to the forecast at query sensor q.
Analysis and proofs	
$\|\mathbf{A}\|_\infty$, $\|Z\|_\infty$	Maximum absolute row sum of $\mathbf{A}$; maximum Euclidean row norm of a grid field Z in the proofs.
B_ζ, L_ζ	Uniform bound and Lipschitz constant of the noiseless code field $\zeta(\boldsymbol{x}) = \tau_\theta(\boldsymbol{u}(\boldsymbol{x}))$.
L_{post}, L_{dec}	Lipschitz constants of the post-encoding map and decoder readout.
L_{spec}, L_{attn}, L_{LN}	Lipschitz bounds for spectral propagation, attention, and LayerNorm.

continued on the next page

Table 3, continued from the previous page.

Symbol	Meaning
m_σ, L_κ	Lower bound on the Gaussian kernel and upper bound on its spatial gradient norm.
$q_{\mathcal{S}}$, c_q, c_-, c_+	Sensor separation radius and constants in the local-bandwidth sampling assumptions.
r_*, $\mathcal{E}_{\mathcal{S}}$	Reference lift and the lift error bound for a sensor layout.
$\mathbf{Z}_{\mathcal{S}}$, $\mathbf{Z}_*$	Discrete and reference post-encoded fields.
$\varepsilon_{\mathrm{LN}}$, d_0	LayerNorm variance stabilizer and attention denominator lower bound.
J_{h_g}, J_*	Discrete grid gather and its continuum reference.
Φ^M_{θ,h_g}, Φ^M_θ	Grid and continuum SRE/DFO stacks.
Q_{h_g}	Endpoint-grid quadrature.
C_{θ,M,h_g}, C_{geo}	Constants controlling operator stability and local interpolation, respectively.
$C_{\mathrm{quad},M}$, $C_{\mathrm{grid},M}$	Grid-to-continuum error constants for the operator stack and full forecasting model, respectively.
δ_χ, δ_q, R_χ	Maximum sensor-coordinate displacement, query-coordinate displacement, and induced prediction discrepancy.
C_β	Constant in the spectral-slope perturbation bound.

C Full Proofs

The analysis follows the lift–condition–evolve–query construction. We first bound the error introduced by sensor sampling, then propagate it through the trained operator, and finally analyze latent-grid refinement and spectral perturbations. Throughout a sampling comparison, the coordinate assignment χ, field history, trained neural maps, latent grid, and common query history are fixed. The period-specific assignments are $\boldsymbol{x}_i^{(p)} = \chi_p(\boldsymbol{s}_i) \in \Omega = [0,1]^2$; coordinate reassignment is quantified separately in Section C.4. Write $\mathcal{S} = \{\boldsymbol{x}_i\}_{i=1}^N$, let $h_{\mathcal{S}}$ be its fill distance, and use the Euclidean channel norm inside every function or grid sup norm. Dark blue labels continuum and reference quantities: the limiting sampling measure μ, smoothed lifts, continuum operators, and reference predictions. Dark orange labels discrete and empirical quantities: sensor layouts, empirical measures, grid quantities, and the predictions computed by the finite model.

C.1 Operator Conventions and Assumptions

Code field and two sampling regimes. The noiseless history field is $\boldsymbol{u}(\boldsymbol{x}) \in \mathbb{R}^{T_h}$ and its temporal encoding is $\zeta(\boldsymbol{x}) = \tau_\theta(\boldsymbol{u}(\boldsymbol{x})) \in \mathbb{R}^d$. Let $\nu_{\mathcal{S}} = N^{-1}\sum_i \delta_{\boldsymbol{x}_i}$ be the empirical sensor measure. For a probability measure ν and bandwidth $\sigma > 0$, the normalized field lift is

$$I_{\nu,\sigma}\zeta(\boldsymbol{x}) = \frac{\int_\Omega \kappa_\sigma(\boldsymbol{x},\boldsymbol{y})\zeta(\boldsymbol{y})\,d\nu(\boldsymbol{y})}{\int_\Omega \kappa_\sigma(\boldsymbol{x},\boldsymbol{y})\,d\nu(\boldsymbol{y})}, \qquad I_{\mathcal{S},\sigma}\zeta = I_{\nu_{\mathcal{S}},\sigma}\zeta. \tag{9}$$

This identity expresses the implemented equal-weight sensor lift as a normalized integral. At fixed learned bandwidth, its reference is the smoothed field $I_{\mu,\sigma}\zeta$ associated with a common limiting sensor measure μ. When the bandwidth decreases with fill distance, its reference is the pointwise code field ζ.

Assumption 1 (Admissible sampling families)**.** Each comparison uses one of the following regimes. In the fixed-bandwidth regime, $\sigma_{\mathrm{enc}} = \sigma > 0$ and $W_1(\nu_{\mathcal{S}},\mu) \to 0$ for a common probability measure μ on Ω. Here $W_1(\nu,\mu)$ is the infimum of $\int \|\boldsymbol{x}-\boldsymbol{y}\|\,d\pi(\boldsymbol{x},\boldsymbol{y})$ over couplings π of ν and μ. In the local-bandwidth regime, the separation radius $q_{\mathcal{S}} = \frac{1}{2}\min_{i\neq j}\|\boldsymbol{x}_i - \boldsymbol{x}_j\|$ satisfies $q_{\mathcal{S}} \geq c_q h_{\mathcal{S}}$, and $c_- h_{\mathcal{S}} \leq \sigma_{\mathrm{enc}} \leq c_+ h_{\mathcal{S}}$ for fixed positive c_q, c_-, c_+, with $h_{\mathcal{S}} \to 0$ along refinement.

These conditions specify respectively convergence of sampling density and convergence of local interpolation neighborhoods. For example, equal-mass cells of μ with diameter at most $Ch_{\mathcal{S}}$, each containing one sensor, give $W_1(\nu_{\mathcal{S}},\mu) \leq Ch_{\mathcal{S}}$ by transporting each cell's mass to its sensor. Thus the fixed-bandwidth result also has a fill-distance rate for such sampling families.

Assumption 2 (Regularity and trained maps)**.** The code field is bounded by B_ζ and is L_ζ-Lipschitz. The coordinate feature map ρ and post-encoding map ψ_θ are Lipschitz. The network has finitely many blocks with finite trained weights and Lipschitz activations. Every LayerNorm uses a positive variance stabilizer $\varepsilon_{\mathrm{LN}}$. The spectral power and denominator stabilizers $\varepsilon_s, \varepsilon_f, \varepsilon_a$ are positive.

For the spectral descriptor, let $\bar{z}_g = d^{-1}\mathbf{1}_d^\top \boldsymbol{z}_g$, $P_j = |\widehat{\bar{z}}(\boldsymbol{k}_j)|^2$, $E_j = P_j + \varepsilon_s$, $k_j = \|\boldsymbol{k}_j\| + \varepsilon_f$, and $x_j = \log k_j - |\mathcal{J}|^{-1}\sum_{r\in\mathcal{J}} \log k_r$. Writing $A = \sum_j E_j$ and $D_\beta = \sum_j x_j^2 + \varepsilon_s$, the statistics in Eq. (3) become

$$\alpha = \frac{\sum_j E_j k_j}{k_{\max} A + \varepsilon_s}, \qquad \beta = -\frac{\sum_j x_j \log E_j}{D_\beta}, \qquad \eta_b = \frac{\sum_{j\in\mathcal{B}_b} E_j}{A + \varepsilon_s}. \tag{10}$$

The centering of $\log E_j$ cancels because $\sum_j x_j = 0$. The state is $\boldsymbol{c} = [\alpha, \mathrm{clip}(\beta, -10, 10), \eta_1, \ldots, \eta_B]^\top$. Effective powers obey $E_j \geq \varepsilon_s$ and $A \geq |\mathcal{J}|\varepsilon_s$.

Assumption 3 (Fixed-design grid refinement)**.** For the grid comparison, the Fourier cutoff M, descriptor frequency set $\mathcal{J}$, band membership, and Fourier normalization are fixed. Grids have n points per axis, $G = n^2$, and $h_g = 1/(n-1)$; retained modes lie below Nyquist for sufficiently large n. The decoder uses either a fixed positive bandwidth or $\sigma_{\mathrm{dec}} \asymp h_g$. All continuum fields reached by the finite stack are bounded and Lipschitz.

This grid comparison concerns the same finite spectral design at increasing resolution. The sensor-sampling comparison keeps the grid and its full spectral design fixed. For $\boldsymbol{a} \in \{0, \ldots, n-1\}^2$, the spatial node is $\boldsymbol{\xi}_{\boldsymbol{a}} = \boldsymbol{a}/(n-1)$ and the FFT phase node is $\boldsymbol{t}_{\boldsymbol{a}} = \boldsymbol{a}/n$. We use the normalized forward transform

$$\widehat{\boldsymbol{z}}_n(\boldsymbol{k}) = \frac{1}{G}\sum_{\boldsymbol{a}} \boldsymbol{z}_{\boldsymbol{a}} e^{-2\pi \mathrm{i}\boldsymbol{k}\cdot\boldsymbol{t}_{\boldsymbol{a}}}, \qquad (\mathcal{P}_{\theta,n}\boldsymbol{z})_{\boldsymbol{a}} = \sum_{\boldsymbol{k}\in\mathcal{K}_M} \mathbf{R}_\theta(\boldsymbol{k})\widehat{\boldsymbol{z}}_n(\boldsymbol{k}) e^{2\pi\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{t}_{\boldsymbol{a}}}. \tag{11}$$

Conjugate symmetry of retained modes and multipliers gives a real output for real input. Section C.5 accounts for the phase-node offset.

C.2 Normalized Kernel Lifting and Gathering

Lemma 1 (Fixed-bandwidth sampling error)**.** *For a fixed $\sigma > 0$, let $m_\sigma = e^{-1/\sigma^2}$ and $L_\kappa = e^{-1/2}/\sigma$. Under Assumption 2,*

$$\|I_{\nu,\sigma}\zeta - I_{\mu,\sigma}\zeta\|_\infty \leq C_{\sigma,\zeta}\, W_1(\nu,\mu), \qquad C_{\sigma,\zeta} = \frac{L_\zeta + 2B_\zeta L_\kappa}{m_\sigma}. \tag{12}$$

Proof. Fix a query $\boldsymbol{x}$. Write $A_\nu = \int \kappa_\sigma(\boldsymbol{x},\boldsymbol{y})\zeta(\boldsymbol{y})\, d\nu(\boldsymbol{y})$ and $D_\nu = \int \kappa_\sigma(\boldsymbol{x},\boldsymbol{y})\, d\nu(\boldsymbol{y})$. The domain diameter is $\sqrt{2}$, so $D_\nu, D_\mu \geq m_\sigma$. The Gaussian derivative satisfies $\|\nabla_{\boldsymbol{y}}\kappa_\sigma(\boldsymbol{x},\boldsymbol{y})\| \leq \max_{r\geq 0} r\sigma^{-2}e^{-r^2/(2\sigma^2)} = L_\kappa$. For any coupling of ν and μ, subtract the paired integrands and apply their Lipschitz bounds. Taking the infimum over couplings gives $\|A_\nu - A_\mu\| \leq (L_\zeta + B_\zeta L_\kappa)W_1(\nu,\mu)$ and $|D_\nu - D_\mu| \leq L_\kappa W_1(\nu,\mu)$. Since $\|A_\mu/D_\mu\| \leq B_\zeta$, the quotient is expanded over the common denominator as

$$\begin{aligned}
\frac{A_\nu}{D_\nu} - \frac{A_\mu}{D_\mu} &= \frac{A_\nu D_\mu - A_\mu D_\nu}{D_\nu D_\mu} \\
&= \frac{(A_\nu - A_\mu)D_\mu + A_\mu(D_\mu - D_\nu)}{D_\nu D_\mu} \\
&= \frac{A_\nu - A_\mu}{D_\nu} + \frac{A_\mu}{D_\mu}\frac{D_\mu - D_\nu}{D_\nu}, \\
\left\|\frac{A_\nu}{D_\nu} - \frac{A_\mu}{D_\mu}\right\| &\leq \frac{\|A_\nu - A_\mu\|}{D_\nu} + \left\|\frac{A_\mu}{D_\mu}\right\|\frac{|D_\mu - D_\nu|}{D_\nu} \\
&\leq \frac{L_\zeta + 2B_\zeta L_\kappa}{m_\sigma} W_1(\nu,\mu).
\end{aligned} \tag{13}$$

The bounds are uniform in $\boldsymbol{x}$, proving the claim. □

Lemma 2 (Local-bandwidth interpolation error)**.** *Under the local-bandwidth regime of Assumption 1, there is a finite C_{geo}, independent of N and $h_{\mathcal{S}}$, such that*

$$\|I_{\mathcal{S},\sigma_{\mathrm{enc}}}\zeta - \zeta\|_\infty \le C_{\mathrm{geo}} L_\zeta h_{\mathcal{S}}. \tag{14}$$

Proof. Let $a_i(\boldsymbol{x})$ be the normalized Gaussian weights. Their positivity and unit sum give

$$\begin{aligned} \|I_{\mathcal{S},\sigma}\zeta(\boldsymbol{x}) - \zeta(\boldsymbol{x})\| &\le \sum_i a_i(\boldsymbol{x})\|\zeta(\boldsymbol{x}_i) - \zeta(\boldsymbol{x})\| \\ &\le L_\zeta \sum_i a_i(\boldsymbol{x})\|\boldsymbol{x}_i - \boldsymbol{x}\|. \end{aligned} \tag{15}$$

Set $h = h_{\mathcal{S}}$. A sensor lies within h of every $\boldsymbol{x}$, so the unnormalized weight sum is at least $c_0 = e^{-1/(2c_-^2)}$. Partition sensors into shells $A_m = \{\boldsymbol{x}_i : mh \le \|\boldsymbol{x}_i - \boldsymbol{x}\| < (m+1)h\}$ for integers $m \ge 0$. The disjoint radius-$q_{\mathcal{S}}$ balls around these sensors fit inside a radius-$(m+1)h + q_{\mathcal{S}}$ ball, so comparing volumes gives

$$|A_m|\,\pi q_{\mathcal{S}}^2 \le \pi\big((m+1)h + q_{\mathcal{S}}\big)^2, \qquad |A_m| \le \left(1 + \frac{(m+1)h}{q_{\mathcal{S}}}\right)^2 \le C_A(m+1)^2, \tag{16}$$

with $C_A = (1 + 1/c_q)^2$, using $q_{\mathcal{S}} \ge c_q h$ and $m + 1 \ge 1$. Consequently,

$$\begin{aligned} \sum_i a_i(\boldsymbol{x})\|\boldsymbol{x}_i - \boldsymbol{x}\| &\le \frac{1}{c_0}\sum_{m\ge 0}|A_m|(m+1)he^{-m^2/(2c_+^2)} \\ &\le \frac{C_A}{c_0} h \sum_{m\ge 0}(m+1)^3 e^{-m^2/(2c_+^2)} = C_{\mathrm{geo}} h. \end{aligned} \tag{17}$$

The Gaussian series converges. Substitution into Eq. (15) completes the proof. □

Post-encoding propagation. For either sampling regime, let r_* be its reference lift and let $\mathcal{E}_{\mathcal{S}}$ denote the corresponding lift error bound:

$$(r_*, \mathcal{E}_{\mathcal{S}}) = \begin{cases} (I_{\mu,\sigma}\zeta, C_{\sigma,\zeta} W_1(\nu_{\mathcal{S}}, \mu)), & \text{fixed bandwidth,} \\ (\zeta, C_{\mathrm{geo}} L_\zeta h_{\mathcal{S}}), & \text{local bandwidth.} \end{cases} \tag{18}$$

Define $\boldsymbol{z}_{\mathcal{S},g} = \psi_\theta([I_{\mathcal{S},\sigma}\zeta(\boldsymbol{\xi}_g)\|\rho(\boldsymbol{\xi}_g)])$ and $\boldsymbol{z}_{*,g} = \psi_\theta([r_*(\boldsymbol{\xi}_g)\|\rho(\boldsymbol{\xi}_g)])$. The identical coordinate features cancel from the input difference, so the L_{post}-Lipschitz map ψ_θ gives $\|\mathbf{Z}_{\mathcal{S}} - \mathbf{Z}_*\|_\infty \le L_{\mathrm{post}}\mathcal{E}_{\mathcal{S}}$.

Lemma 3 (Decoder gather error)**.** *The normalized Gaussian gather J_{h_g} is nonexpansive from the grid sup norm to the query norm. For a bounded Lipschitz field v, it satisfies*

$$\left\|J_{h_g}(v|_{\Xi_{h_g}})(\boldsymbol{x}_q) - J_* v(\boldsymbol{x}_q)\right\| \le C_{\mathrm{dec},v} h_g, \qquad J_* v = \begin{cases} I_{\lambda,\sigma_{\mathrm{dec}}} v, & \sigma_{\mathrm{dec}} > 0 \text{ fixed}, \\ v, & \sigma_{\mathrm{dec}} \asymp h_g, \end{cases} \tag{19}$$

where λ is Lebesgue probability measure on the unit square.

Proof. For arbitrary grid fields V, V', positivity of b_{qg} and $\sum_g b_{qg} = 1$ give $\|\sum_g b_{qg}(\boldsymbol{v}_g - \boldsymbol{v}'_g)\| \le \|V - V'\|_\infty$. For fixed bandwidth, partition Ω into n^2 equal squares and couple each square to its associated endpoint-grid node. For node $\boldsymbol{\xi}_{\boldsymbol{a}} = \boldsymbol{a}/(n-1)$ and cell $C_{\boldsymbol{a}} = \prod_{j=1}^2 [a_j/n, (a_j + 1)/n]$, each axis contributes at most $1/n \le h_g$, so each transport distance is at most $\sqrt{2}h_g$ and $W_1(G^{-1}\sum_g \delta_{\boldsymbol{\xi}_g}, \lambda) \le \sqrt{2}h_g$. Lemma 1 with $\zeta = v$ proves the first case. For $\sigma_{\mathrm{dec}} \asymp h_g$, the grid is quasi-uniform and the first-moment proof gives

$$\begin{aligned} \left\|\sum_g b_{qg} v(\boldsymbol{\xi}_g) - v(\boldsymbol{x}_q)\right\| &\le L_v \sum_g b_{qg}\|\boldsymbol{\xi}_g - \boldsymbol{x}_q\| \\ &\le C_{\mathrm{geo}} L_v h_g, \end{aligned} \tag{20}$$

with the grid geometry absorbed into C_{geo}. □

C.3 Stability of the Trained Field Operator

Spectral descriptor and conditioning. A normalized Fourier coefficient is a bounded linear map with sup-norm gain at most one. The power difference satisfies $\big||a|^2-|b|^2\big| \le (|a|+|b|)|a-b|$. On bounded fields, the stabilized logarithms and ratios in Eq. (10) therefore have finite Lipschitz constants; their denominators are bounded below by positive stabilizers. Clipping is nonexpansive. Denote the resulting descriptor constant by L_c. In block ℓ, $\|\boldsymbol{\gamma}_\ell\|_\infty \le 0.1$ and $(\boldsymbol{\gamma}_\ell, \boldsymbol{\delta}_\ell)$ is Lipschitz in $\boldsymbol{c}$. For fields bounded by B_ℓ, the modulation $\mathbf{X}_\ell^{\mathrm{c}} = \mathbf{X}_\ell \odot (\mathbf{1}_G(\mathbf{1}_d+\boldsymbol{\gamma}_\ell)^\top) + \mathbf{1}_G\boldsymbol{\delta}_\ell^\top$ expands into three terms,

$$\mathbf{X}_\ell^{\mathrm{c}} - \mathbf{X}_\ell'^{\mathrm{c}} = (\mathbf{X}_\ell - \mathbf{X}_\ell') \odot (\mathbf{1}_G(\mathbf{1}_d+\boldsymbol{\gamma}_\ell)^\top) + \mathbf{X}_\ell' \odot (\mathbf{1}_G(\boldsymbol{\gamma}_\ell - \boldsymbol{\gamma}_\ell')^\top) + \mathbf{1}_G(\boldsymbol{\delta}_\ell - \boldsymbol{\delta}_\ell')^\top,$$

$$\begin{aligned}\|X_\ell^{\mathrm{c}} - X_\ell'^{\mathrm{c}}\|_\infty &\le 1.1\|X_\ell - X_\ell'\|_\infty + B_\ell\|\boldsymbol{\gamma}_\ell(\boldsymbol{c}) - \boldsymbol{\gamma}_\ell(\boldsymbol{c}')\| + \|\boldsymbol{\delta}_\ell(\boldsymbol{c}) - \boldsymbol{\delta}_\ell(\boldsymbol{c}')\| \\ &\le 1.1\|X_\ell - X_\ell'\|_\infty + (B_\ell L_{\gamma,\ell} + L_{\delta,\ell})\|\boldsymbol{c} - \boldsymbol{c}'\|,\end{aligned} \tag{21}$$

where the first line uses $\|\mathbf{1}_d + \boldsymbol{\gamma}_\ell\|_\infty \le 1.1$ and the field bound.

Lemma 4 (Normalized linear attention)**.** *On bounded feature sets, the attention map in Eq.* (6) *has a Lipschitz constant independent of* G.

Proof. Use column vectors $\boldsymbol{q}_g, \boldsymbol{k}_g \in \mathbb{R}^{d_a}$ and $\boldsymbol{v}_g \in \mathbb{R}^d$, and form the matrices $\mathbf{Q} = [\boldsymbol{q}_1, \dots, \boldsymbol{q}_G]^\top \in \mathbb{R}^{G\times d_a}$, $\mathbf{K} = [\boldsymbol{k}_1, \dots, \boldsymbol{k}_G]^\top \in \mathbb{R}^{G\times d_a}$, and $\mathbf{V} = [\boldsymbol{v}_1, \dots, \boldsymbol{v}_G]^\top \in \mathbb{R}^{G\times d}$. The attention features before the output projection have the exact matrix form

$$\begin{aligned}&\mathbf{T} = \mathrm{diag}(\mathbf{Q}\mathbf{K}^\top\mathbf{1}_G + \varepsilon_a\mathbf{1}_G)^{-1}\mathbf{Q}(\mathbf{K}^\top\mathbf{V}), \\ &\boldsymbol{t}_g^\top = \frac{\boldsymbol{q}_g^\top\overline{\mathbf{B}}^\top}{\boldsymbol{q}_g^\top\overline{\boldsymbol{k}} + \varepsilon_a/G}, \qquad \overline{\mathbf{B}} = \frac{1}{G}\mathbf{V}^\top\mathbf{K} = \frac{1}{G}\sum_j \boldsymbol{v}_j\boldsymbol{k}_j^\top, \quad \overline{\boldsymbol{k}} = \frac{1}{G}\mathbf{K}^\top\mathbf{1}_G = \frac{1}{G}\sum_j \boldsymbol{k}_j,\end{aligned} \tag{22}$$

where $\boldsymbol{t}_g^\top$ is row g of $\mathbf{T}$. The first line is an algebraic identity that reads entrywise. Its diagonal factor collects the row sums $[\mathbf{Q}\mathbf{K}^\top\mathbf{1}_G]_g = \sum_j \boldsymbol{q}_g^\top\boldsymbol{k}_j$, while $\mathbf{Q}(\mathbf{K}^\top\mathbf{V})$ has row g

$$\boldsymbol{q}_g^\top\Big(\sum_j \boldsymbol{k}_j\boldsymbol{v}_j^\top\Big) = \sum_j (\boldsymbol{q}_g^\top\boldsymbol{k}_j)\boldsymbol{v}_j^\top, \tag{23}$$

so row g of $\mathbf{T}$ reproduces the fraction in Eq. (6). The second line evaluates the same row through the two shared aggregates, because

$$\overline{\mathbf{B}}\boldsymbol{q}_g = \frac{1}{G}\sum_j (\boldsymbol{q}_g^\top\boldsymbol{k}_j)\boldsymbol{v}_j, \qquad \boldsymbol{q}_g^\top\overline{\boldsymbol{k}} + \frac{\varepsilon_a}{G} = \frac{1}{G}\Big(\sum_j \boldsymbol{q}_g^\top\boldsymbol{k}_j + \varepsilon_a\Big) \tag{24}$$

factor a common G^{-1} from numerator and denominator. Bounded affine preactivations and $\phi(t) = \mathrm{ELU}(t) + 1 > 0$ give a common componentwise interval $[m_\phi, B_\phi]$ with $m_\phi > 0$, so every entry of $\boldsymbol{q}_g$ and $\boldsymbol{k}_j$ is at least m_ϕ and

$$\boldsymbol{q}_g^\top\boldsymbol{k}_j = \sum_{a=1}^{d_a} q_{g,a}k_{j,a} \ge d_a m_\phi^2. \tag{25}$$

Thus every denominator is at least $d_0 = d_a m_\phi^2$, independently of G.

Let B_q, B_k, B_v bound the projected vector norms and let L_q, L_k, L_v be their Lipschitz constants. If $e = \|X - X'\|_\infty$, one outer product splits as $\boldsymbol{v}_j\boldsymbol{k}_j^\top - \boldsymbol{v}_j'\boldsymbol{k}_j'^\top = (\boldsymbol{v}_j - \boldsymbol{v}_j')\boldsymbol{k}_j^\top + \boldsymbol{v}_j'(\boldsymbol{k}_j - \boldsymbol{k}_j')^\top$, so

$$\begin{aligned}\|\overline{\mathbf{B}} - \overline{\mathbf{B}}'\| &\le \frac{1}{G}\sum_j \|(\boldsymbol{v}_j - \boldsymbol{v}_j')\boldsymbol{k}_j^\top + \boldsymbol{v}_j'(\boldsymbol{k}_j - \boldsymbol{k}_j')^\top\| \\ &\le (B_kL_v + B_vL_k)e, \qquad \|\overline{\boldsymbol{k}} - \overline{\boldsymbol{k}}'\| \le L_k e.\end{aligned} \tag{26}$$

For $N_g = \overline{\mathbf{B}}\boldsymbol{q}_g$ and $D_g = \boldsymbol{q}_g^\top\overline{\boldsymbol{k}} + \varepsilon_a/G$, this implies $\|N_g\| \le B_N = B_vB_kB_q$, $\|N_g - N_g'\| \le L_N e$, and $|D_g - D_g'| \le L_D e$, with $L_N = B_q(B_kL_v + B_vL_k) + B_vB_kL_q$ and $L_D = B_kL_q + B_qL_k$. The

quotient expansion yields

$$\left\| \frac{N_g}{D_g} - \frac{N'_g}{D'_g} \right\| \le \frac{\|N_g - N'_g\|}{D_g} + \frac{\|N'_g\|\,|D'_g - D_g|}{D_g D'_g} \le \left(\frac{L_N}{d_0} + \frac{B_N L_D}{d_0^2} \right) e. \tag{27}$$

The affine output map contributes the norm L_O of its linear part. Hence $L_{\mathrm{attn}} = L_O(L_N/d_0 + B_N L_D/d_0^2)$ is independent of G. □

Residual normalization. The retained Fourier path has gain $L_{\mathrm{spec}} \le \sum_{\boldsymbol{k}\in\mathcal{K}_M} \|\mathbf{R}_\theta(\boldsymbol{k})\|$ by the triangle inequality over modes. For LayerNorm, write the centering projection $\mathbf{C} = \mathbf{I} - \mathbf{1}_d\mathbf{1}_d^\top/d$, $s = (\|\mathbf{C}\boldsymbol{v}\|^2/d + \varepsilon_{\mathrm{LN}})^{1/2}$, and let $\boldsymbol{a}, \boldsymbol{b}$ be its learned scale and bias. Since $\mathbf{C}$ is symmetric and idempotent, $\partial s/\partial \boldsymbol{v} = \mathbf{C}\boldsymbol{v}/(ds)$, and differentiating the quotient gives

$$\begin{aligned} \mathrm{LN}(\boldsymbol{v}) &= \mathrm{diag}(\boldsymbol{a})\frac{\mathbf{C}\boldsymbol{v}}{s} + \boldsymbol{b}, \\ D\,\mathrm{LN}(\boldsymbol{v}) &= \mathrm{diag}(\boldsymbol{a})\left(\frac{\mathbf{C}}{s} - \frac{(\mathbf{C}\boldsymbol{v})(\mathbf{C}\boldsymbol{v})^\top}{ds^3} \right), \\ \|D\,\mathrm{LN}(\boldsymbol{v})\| &\le \frac{\|\boldsymbol{a}\|_\infty}{\sqrt{\varepsilon_{\mathrm{LN}}}}. \end{aligned} \tag{28}$$

The bracket is symmetric with the eigenpairs $\mathbf{C}\boldsymbol{v} \mapsto \varepsilon_{\mathrm{LN}}/s^3$ and $\boldsymbol{w} \mapsto 1/s$ for $\boldsymbol{w} \perp \{\mathbf{1}_d, \mathbf{C}\boldsymbol{v}\}$, and it annihilates $\mathbf{1}_d$. The first eigenvalue follows from $\|\mathbf{C}\boldsymbol{v}\|^2 = d(s^2 - \varepsilon_{\mathrm{LN}})$, and $s \ge \sqrt{\varepsilon_{\mathrm{LN}}}$ shows that $1/s$ is the larger one, so the spectral norm of the bracket is $1/s$. This establishes the stated bound and also bounds every block output on bounded inputs.

Lemma 5 (Finite stack with shared spectral state)**.** *For a fixed query, let $\Psi_{\theta,h_g,M}$ map the post-encoded field to the field-path forecast through all L operator blocks and the decoder. Under Assumption 2,*

$$\|\Psi_{\theta,h_g,M}(Z) - \Psi_{\theta,h_g,M}(Z')\| \le C_{\theta,M,h_g}\|Z - Z'\|_\infty. \tag{29}$$

Proof. Set $X_0 = Z$, $X'_0 = Z'$, and $e_\ell = \|X_\ell - X'_\ell\|_\infty$. Every block uses the initial descriptor, so $\|\mathbf{c}(Z) - \mathbf{c}(Z')\| \le L_c e_0$ throughout the stack. Combining Eq. (21) with the unconditioned residual connection gives the scalar recursion

$$\begin{aligned} e_{\ell+1} &\le L_{\mathrm{LN},\ell}\,[e_\ell + (L_{\mathrm{spec},\ell} + L_{\mathrm{attn},\ell})\{1.1e_\ell + (B_\ell L_{\gamma,\ell} + L_{\delta,\ell})L_c e_0\}] \\ &= a_\ell e_\ell + b_\ell L_c e_0. \end{aligned} \tag{30}$$

Unrolling it gives $e_1 \le a_0 e_0 + b_0 L_c e_0$, $e_2 \le a_1 a_0 e_0 + (a_1 b_0 + b_1)L_c e_0$, and in general

$$e_L \le \left[\prod_{\ell=0}^{L-1} a_\ell + L_c \sum_{j=0}^{L-1} b_j \prod_{\ell=j+1}^{L-1} a_\ell \right] e_0. \tag{31}$$

Here $a_\ell = L_{\mathrm{LN},\ell}[1 + 1.1(L_{\mathrm{spec},\ell} + L_{\mathrm{attn},\ell})]$, $b_\ell = L_{\mathrm{LN},\ell}(L_{\mathrm{spec},\ell} + L_{\mathrm{attn},\ell})(B_\ell L_{\gamma,\ell} + L_{\delta,\ell})$, and an empty product equals one. The decoder gather is nonexpansive. Multiplying the bracket by the final normalization/readout constant L_{dec} defines C_{θ,M,h_g}. For fixed spectral design, these constants can also be chosen uniformly over sufficiently fine grids. □

C.4 Sensor-Discretization Consistency

Theorem 1 (Sensor-discretization consistency)**.** *Fix the coordinate representation, field history, trained neural maps, latent grid, spectral design, and common query. Under Assumptions 1–2, the reference lift r_* in Eq. (18) defines a sampling-independent prediction $\overline{\mathcal{G}}_\theta^{h_g,M}(\boldsymbol{x}_q)$ satisfying*

$$\left\| \mathcal{G}_\theta^{\mathcal{S},h_g,M}(\boldsymbol{x}_q) - \overline{\mathcal{G}}_\theta^{h_g,M}(\boldsymbol{x}_q) \right\| \le C_{\theta,M,h_g} L_{\mathrm{post}} \mathcal{E}_{\mathcal{S}}. \tag{32}$$

At fixed bandwidth the error is $O(W_1(\nu_{\mathcal{S}}, \mu))$; under local-bandwidth refinement it is $O(h_{\mathcal{S}})$.

Proof. The common local skip is $\boldsymbol{s}_q = \mathbf{W}_{\text{skip}}\boldsymbol{u}_q + \boldsymbol{b}_{\text{skip}}$. With the post-encoded fields defined after Eq. (18), the full predictions factor as

$$\begin{aligned} \mathcal{G}_\theta^{\mathcal{S},h_g,M}(\boldsymbol{x}_q) &= \Psi_{\theta,h_g,M}(\mathbf{Z}_{\mathcal{S}}) + \boldsymbol{s}_q, \\ \overline{\mathcal{G}}_\theta^{h_g,M}(\boldsymbol{x}_q) &= \Psi_{\theta,h_g,M}(\mathbf{Z}_*) + \boldsymbol{s}_q, \\ \|\mathcal{G}_\theta^{\mathcal{S},h_g,M}(\boldsymbol{x}_q) - \overline{\mathcal{G}}_\theta^{h_g,M}(\boldsymbol{x}_q)\| &\le C_{\theta,M,h_g}\|\mathbf{Z}_{\mathcal{S}} - \mathbf{Z}_*\|_\infty \\ &\le C_{\theta,M,h_g} L_{\text{post}} \mathcal{E}_{\mathcal{S}}. \end{aligned} \tag{33}$$

The reference uses the same trained maps, common query, and reference field for every layout in its sampling family. The respective sampling assumptions make $\mathcal{E}_{\mathcal{S}}$ converge to zero. □

Corollary 1 (Expansion stability)**.** *Let two layouts $\mathcal{S}, \mathcal{S}'$ share a sampling regime and its reference field, as well as the fixed quantities of Theorem 1. Then*

$$\begin{aligned} \|\mathcal{G}_\theta^{\mathcal{S}',h_g,M}(\boldsymbol{x}_q) - \mathcal{G}_\theta^{\mathcal{S},h_g,M}(\boldsymbol{x}_q)\| &\le \|\mathcal{G}_\theta^{\mathcal{S}',h_g,M}(\boldsymbol{x}_q) - \overline{\mathcal{G}}_\theta^{h_g,M}(\boldsymbol{x}_q)\| + \|\overline{\mathcal{G}}_\theta^{h_g,M}(\boldsymbol{x}_q) - \mathcal{G}_\theta^{\mathcal{S},h_g,M}(\boldsymbol{x}_q)\| \\ &\le C_{\theta,M,h_g} L_{\text{post}}(\mathcal{E}_{\mathcal{S}'} + \mathcal{E}_{\mathcal{S}}). \end{aligned} \tag{34}$$

In particular, local-bandwidth refinement gives $C_{\text{samp}}(h_{\mathcal{S}'} + h_{\mathcal{S}})$, where $C_{\text{samp}} = C_{\theta,M,h_g} L_{\text{post}} C_{\text{geo}} L_\zeta$. For nested layouts, $h_{\mathcal{S}'} \le h_{\mathcal{S}}$.

Coordinate reassignment. Fix histories, bandwidths, neural maps, and the grid for the same sensor identities V. Let χ, χ' be two assignments and set $\delta_\chi = \max_{i\in V}\|\chi'(\boldsymbol{s}_i) - \chi(\boldsymbol{s}_i)\|$ and $\delta_q = \|\chi'(\boldsymbol{s}_q) - \chi(\boldsymbol{s}_q)\|$. Interpolate the coordinates linearly between the assignments. For encoder logits $\ell_i = -\|\boldsymbol{\xi}_g - \boldsymbol{x}_i\|^2/(2\sigma_{\text{enc}}^2)$, $|\dot{\ell}_i| \le \sqrt{2}\delta_\chi/\sigma_{\text{enc}}^2$. Differentiating the normalized weights gives $\dot{a}_i = a_i(\dot{\ell}_i - \sum_j a_j\dot{\ell}_j)$ and thus $\sum_i |\dot{a}_i| \le 2\sqrt{2}\delta_\chi/\sigma_{\text{enc}}^2$. Applying the same calculation to decoder weights and integrating along the coordinate path yields the reassignment bound

$$R_\chi \le \frac{2\sqrt{2}C_{\theta,M,h_g}L_{\text{post}}B_\zeta}{\sigma_{\text{enc}}^2}\delta_\chi + \frac{2\sqrt{2}L_{\text{dec}}B_H}{\sigma_{\text{dec}}^2}\delta_q, \tag{35}$$

where B_H bounds the evolved field along the comparison and R_χ is the norm of the difference between the two full predictions. For an expansion $V \subseteq V'$, suppose the layouts under χ' share the sampling regime and reference field of Theorem 1. Insert the prediction on V under χ' to obtain

$$\|\mathcal{G}_\theta(\mathcal{D}_{\chi'}(V'))(\chi'(\boldsymbol{s}_q)) - \mathcal{G}_\theta(\mathcal{D}_\chi(V))(\chi(\boldsymbol{s}_q))\| \le C_{\theta,M,h_g}L_{\text{post}}(\mathcal{E}_{\mathcal{S}_{\chi'}(V')} + \mathcal{E}_{\mathcal{S}_{\chi'}(V)}) + R_\chi. \tag{36}$$

Here $\mathcal{D}_\chi(V) = \{(\chi(\boldsymbol{s}_i), \boldsymbol{u}_i) : i \in V\}$ and $\mathcal{S}_\chi(V) = \{\chi(\boldsymbol{s}_i) : i \in V\}$. The two sampling errors are evaluated in the new representation. This decomposition separates the contributions of sensor sampling and coordinate motion.

C.5 Fixed-Design Grid-to-Continuum Consistency

Lemma 6 (Endpoint-grid quadrature)**.** *For an L_f-Lipschitz function $f : \Omega \to \mathbb{R}^m$, the endpoint-grid quadrature $Q_{h_g}f = G^{-1}\sum_g f(\boldsymbol{\xi}_g)$ satisfies $\|Q_{h_g}f - \int_\Omega f(\boldsymbol{x})\,d\boldsymbol{x}\| \le \sqrt{2}L_f h_g$.*

Proof. Associate node $\boldsymbol{\xi}_{\boldsymbol{a}} = \boldsymbol{a}/(n-1)$ with the equal-mass cell $C_{\boldsymbol{a}} = \prod_{j=1}^2[a_j/n, (a_j+1)/n]$; shared boundaries have zero measure. Then $|C_{\boldsymbol{a}}| = 1/G$ and $\sup_{\boldsymbol{x}\in C_{\boldsymbol{a}}}\|\boldsymbol{x} - \boldsymbol{\xi}_{\boldsymbol{a}}\| \le \sqrt{2}h_g$. Consequently,

$$\begin{aligned} \left\| Q_{h_g}f - \int_\Omega f(\boldsymbol{x})\,d\boldsymbol{x} \right\| &\le \sum_{\boldsymbol{a}} \int_{C_{\boldsymbol{a}}} \|f(\boldsymbol{\xi}_{\boldsymbol{a}}) - f(\boldsymbol{x})\|\,d\boldsymbol{x} \\ &\le \sqrt{2}L_f h_g \sum_{\boldsymbol{a}} |C_{\boldsymbol{a}}| = \sqrt{2}L_f h_g. \end{aligned} \tag{37}$$

□

Lemma 7 (Consistency of spectral conditioning and the operator stack)**.** *Let Φ_{θ,h_g}^M be the grid* SRE/DFO *stack and let Φ_θ^M be its continuum counterpart using normalized Fourier integrals, the*

same finite modes and descriptor bins, and normalized attention integrals. Under Assumptions 2–3, for a bounded Lipschitz input z,

$$\max_g \|\Phi^M_{\theta,h_g}(z|_{\Xi_{h_g}})_g - \Phi^M_\theta(z)(\boldsymbol{\xi}_g)\| \le C_{\mathrm{quad},M} h_g. \tag{38}$$

Proof. We track the Fourier, conditioning, and attention errors before propagating them through the stack.

Fourier coefficients and phases. Let w be bounded by B_w and L_w-Lipschitz, and let $\widehat{w}(\boldsymbol{k}) = \int_\Omega w(\boldsymbol{x}) e^{-2\pi \mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}}\, d\boldsymbol{x}$. The integrand has Lipschitz constant at most $L_w + 2\pi\|\boldsymbol{k}\|B_w$, and the phase factor obeys $|e^{\mathrm{i}x} - e^{\mathrm{i}y}| \le |x-y|$ with $\|\boldsymbol{t}_{\boldsymbol{a}} - \boldsymbol{\xi}_{\boldsymbol{a}}\| \le \sqrt{2}h_g$. Splitting the grid sum from its continuum integral at the spatial nodes therefore gives

$$\begin{aligned}
\|\widehat{w}_n(\boldsymbol{k}) - \widehat{w}(\boldsymbol{k})\| &\le \frac{1}{G}\sum_{\boldsymbol{a}} \|w(\boldsymbol{\xi}_{\boldsymbol{a}})\| \left|e^{-2\pi\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{t}_{\boldsymbol{a}}} - e^{-2\pi\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{\xi}_{\boldsymbol{a}}}\right| \\
&\quad + \left\| Q_{h_g}(we^{-2\pi\mathrm{i}\boldsymbol{k}\cdot(\cdot)}) - \int_\Omega we^{-2\pi\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}}\, d\boldsymbol{x} \right\| \\
&\le 2\sqrt{2}\pi\|\boldsymbol{k}\|B_w h_g + \sqrt{2}(L_w + 2\pi\|\boldsymbol{k}\|B_w)h_g \\
&\le \sqrt{2}(L_w + 4\pi\|\boldsymbol{k}\|B_w)h_g.
\end{aligned} \tag{39}$$

The inverse transform has the corresponding phase correction. Expanding mode by mode gives

$$\begin{aligned}
\|(\mathcal{P}_{\theta,n}w)_g - \mathcal{P}_\theta w(\boldsymbol{\xi}_g)\| &\le \sum_{\boldsymbol{k}\in\mathcal{K}_M} \|\mathbf{R}_\theta(\boldsymbol{k})\| \left[\|\widehat{w}_n(\boldsymbol{k}) - \widehat{w}(\boldsymbol{k})\| + \|\widehat{w}(\boldsymbol{k})\| |e^{2\pi\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{t}_g} - e^{2\pi\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{\xi}_g}|\right] \\
&\le C_{\mathrm{spec},M} h_g.
\end{aligned} \tag{40}$$

Both coefficient and evaluation-phase errors therefore have first-order bounds for the same finite retained modes.

Descriptor and common conditioning state. Equation (39) applies to the fixed bins in $\mathcal{J}$. The power difference bound, positive stabilizers, and nonexpansive clipping yield $\|\boldsymbol{c}_{h_g} - \boldsymbol{c}_*\| \le C_c h_g$. Every block receives this same descriptor discrepancy. Its contribution to FiLM is bounded by Eq. (21).

Normalized attention integrals. For a continuum conditioned field, let $\mathbf{B}_* = \int_\Omega \boldsymbol{v}(\boldsymbol{y})\boldsymbol{k}(\boldsymbol{y})^\top d\boldsymbol{y}$ and $\boldsymbol{k}_* = \int_\Omega \boldsymbol{k}(\boldsymbol{y})\, d\boldsymbol{y}$. Its attention output before projection is $\mathbf{T}_*(\boldsymbol{x}) = \mathbf{B}_*\boldsymbol{q}(\boldsymbol{x})/(\boldsymbol{q}(\boldsymbol{x})^\top \boldsymbol{k}_*)$. The integrands are bounded and Lipschitz, so endpoint-grid quadrature gives $\|\overline{\mathbf{B}} - \mathbf{B}_*\| + \|\overline{\boldsymbol{k}} - \boldsymbol{k}_*\| = O(h_g)$. Both normalized denominators have lower bound d_0. The grid denominator additionally contains ε_a/G, giving

$$\left\| \frac{\overline{\mathbf{B}}\boldsymbol{q}_g}{\boldsymbol{q}_g^\top \overline{\boldsymbol{k}} + \varepsilon_a/G} - \frac{\mathbf{B}_*\boldsymbol{q}_g}{\boldsymbol{q}_g^\top \boldsymbol{k}_*} \right\| \le \frac{B_q\|\overline{\mathbf{B}} - \mathbf{B}_*\|}{d_0} + \frac{B_N}{d_0^2}\left(B_q\|\overline{\boldsymbol{k}} - \boldsymbol{k}_*\| + \frac{\varepsilon_a}{G}\right) = O(h_g). \tag{41}$$

Here $G^{-1} = n^{-2} = O(h_g^2)$. The affine output projection preserves the rate. Changes in the conditioned input contribute another first-order term by Lemma 4.

Propagation through all blocks. Let e_ℓ be the maximum difference between grid and continuum samples after block ℓ, with $e_0 = 0$. The preceding estimates and residual normalization give $e_{\ell+1} \le a_\ell e_\ell + d_\ell h_g$, where d_ℓ includes the shared descriptor and quadrature errors. Expanding the recursion gives $e_L \le h_g \sum_{j=0}^{L-1} d_j \prod_{\ell=j+1}^{L-1} a_\ell$. The finite sum defines $C_{\mathrm{quad},M}$ and proves the lemma. □

Theorem 2 (Fixed-design grid consistency)**.** *Under Assumptions 1–3, let $\mathcal{G}^M_\theta$ use the continuum spectral/attention stack and the decoder reference J_* in Lemma 3. For either sampling reference in Eq.* (18)*,*

$$\begin{aligned}
\|\overline{\mathcal{G}}^{h_g,M}_\theta(\boldsymbol{x}_q) - \mathcal{G}^M_\theta(\boldsymbol{x}_q)\| &\le C_{\mathrm{grid},M} h_g, \\
\|\mathcal{G}^{\mathcal{S},h_g,M}_\theta(\boldsymbol{x}_q) - \mathcal{G}^M_\theta(\boldsymbol{x}_q)\| &\le C_{\theta,M,h_g} L_{\mathrm{post}} \mathcal{E}_{\mathcal{S}} + C_{\mathrm{grid},M} h_g.
\end{aligned} \tag{42}$$

Proof. The reference post-encoding field $z_*(\boldsymbol{x}) = \psi_\theta([r_*(\boldsymbol{x})\|\rho(\boldsymbol{x})])$ is bounded and Lipschitz. For fixed encoder bandwidth, this follows by differentiating its positive Gaussian numerator and denominator; for local bandwidth it follows from the assumed regularity of ζ. Let $H_{h_g} = \Phi^M_{\theta,h_g}(z_*|_{\Xi_{h_g}})$

and $H_* = \Phi_\theta^M(z_*)$. Lemma 7 gives $\|H_{h_g} - H_*|_{\Xi_{h_g}}\|_\infty \le C_{\mathrm{quad},M} h_g$. The common local skip cancels, while adding and subtracting the gather of H_* yields

$$\begin{aligned}\|\overline{\mathcal{G}}_\theta^{h_g,M}(\boldsymbol{x}_q) - \mathcal{G}_\theta^M(\boldsymbol{x}_q)\| &\le L_{\mathrm{dec}}\|J_{h_g}(H_{h_g} - H_*|_{\Xi_{h_g}})(\boldsymbol{x}_q)\| + L_{\mathrm{dec}}\|J_{h_g}(H_*|_{\Xi_{h_g}})(\boldsymbol{x}_q) - J_* H_*(\boldsymbol{x}_q)\| \\ &\le L_{\mathrm{dec}}(C_{\mathrm{quad},M} + C_{\mathrm{dec},H_*})h_g.\end{aligned} \tag{43}$$

This defines $C_{\mathrm{grid},M}$. Adding the sensor bound in Theorem 1 proves the joint estimate. □

C.6 Perturbation Stability of the Spectral State

Theorem 3 (Stabilized spectral descriptor). *Let $P_j, P'_j \ge 0$ be spectra on the same frequency set, with $E_j = P_j + \varepsilon_s$, $E'_j = P'_j + \varepsilon_s$, and $\Delta_j = E'_j - E_j$. If $\eta = \|\Delta\|_1/\|E\|_1 < 1$, then*

$$|\alpha(P') - \alpha(P)| \le \frac{\eta}{1-\eta}, \qquad |\eta_b(P') - \eta_b(P)| \le \frac{\eta}{1-\eta}. \tag{44}$$

If also $\max_j |\Delta_j|/E_j \le \eta_\infty < 1$, the clipped slope obeys

$$|\widetilde{\beta}(P') - \widetilde{\beta}(P)| \le C_\beta \frac{\eta_\infty}{1-\eta_\infty}, \qquad C_\beta = \frac{\sqrt{|\mathcal{J}|}\,\|x\|_2}{D_\beta}. \tag{45}$$

Proof. Write $A = \sum_j E_j$ and $D = \sum_j \Delta_j$, so $A + D \ge (1-\eta)A$. Both the centroid and band ratios have the form $F(E) = \sum_j E_j s_j/(A+\tau)$ with $s_j \in [0,1]$ and $\tau > 0$: for the centroid use $s_j = k_j/k_{\max}$ and $\tau = \varepsilon_s/k_{\max}$; for band b use $s_j = \mathbf{1}\{j \in \mathcal{B}_b\}$ and $\tau = \varepsilon_s$. Thus $F(E) \in [0,1]$, and expansion of the ratio gives

$$\begin{aligned}F(E+\Delta) - F(E) &= \frac{\sum_j \Delta_j s_j - F(E)\sum_j \Delta_j}{A + D + \tau} \\ &= \frac{\sum_j \Delta_j (s_j - F(E))}{A + D + \tau}, \\ |F(E+\Delta) - F(E)| &\le \frac{\|\Delta\|_1}{(1-\eta)A + \tau} \le \frac{\eta}{1-\eta}.\end{aligned} \tag{46}$$

This proves both ratio bounds for signed power changes. For the slope, let $\delta_j = \Delta_j/E_j$. Since $\sum_j x_j = 0$, the centered log-power difference reduces to

$$\begin{aligned}\beta(P') - \beta(P) &= -\frac{1}{D_\beta}\sum_j x_j \left[\log E'_j - \log E_j - \frac{1}{|\mathcal{J}|}\sum_r (\log E'_r - \log E_r)\right] \\ &= -\frac{1}{D_\beta}\sum_j x_j \log(1+\delta_j), \\ |\beta(P') - \beta(P)| &\le \frac{\|x\|_2\,\|\log(1+\delta)\|_2}{D_\beta} \le C_\beta \frac{\eta_\infty}{1-\eta_\infty}.\end{aligned} \tag{47}$$

The last inequality uses $|\log(1+\delta_j)| \le -\log(1-\eta_\infty) \le \eta_\infty/(1-\eta_\infty)$. Clipping to $[-10,10]$ is 1-Lipschitz, completing the result. □

Amplitude dependence. For positive raw powers, omit the additive power and ratio stabilizers and write the resulting statistics as $\alpha_0, \eta_{b,0}, \beta_0$, keeping D_β fixed. Multiplying the latent field by $a > 0$ scales raw power by a^2. Common factors cancel from the idealized centroid and band ratios, while the centered slope satisfies

$$\beta_0(a^2 P) = -\frac{\sum_j x_j(2\log a + \log P_j)}{D_\beta} = -\frac{2\log a \sum_j x_j}{D_\beta} - \frac{\sum_j x_j \log P_j}{D_\beta} = \beta_0(P), \tag{48}$$

Thus the idealized state measures relative spectral shape. In the stabilized state, $E'_j = a^2 P_j + \varepsilon_s$ gives a continuous amplitude dependence, which approaches the idealized statistics as power grows relative to the stabilizers. General additive field changes produce signed per-bin power perturbations through their Fourier cross terms; Theorem 3 quantifies the corresponding descriptor change.

D EXPERIMENTAL SETUP DETAILS

Architecture. We evaluate three model scales with hidden widths $d = 32$ (Small), $d = 64$ (Medium), and $d = 128$ (Large). Within each dataset, the three scales differ only in hidden width. All configurations use a latent grid with $n_g = 32$ points per axis, a Fourier cutoff of $M = 4$, attention width $d_a = 32$, and depthwise spectral mixing. The temporal encoder consists of a linear projection followed by layer normalization and GELU, while CQD applies layer normalization followed by a linear readout. The coordinate encoding uses 16 random frequencies, whose sine and cosine components yield $d_\rho = 32$ features. We use $L = 2$ operator layers for PEMS-Stream and CA-Stream and $L = 1$ for AIR-Stream, with $B = 4$ spectral bands for PEMS-Stream and $B = 3$ for CA-Stream and AIR-Stream. The field readout is combined with a learnable local-history skip connection shared across sensors.

Coordinate assignment. We use three types of static descriptors: geographic coordinates, numerical attributes, and categorical attributes. Numerical attributes are median-imputed and standardized, while categorical attributes are one-hot encoded. The processed numerical and categorical attributes are jointly reduced to two dimensions using PCA. Preprocessing statistics and the PCA projection are estimated from the first period and reused in subsequent periods. Within each period, geographic coordinates and PCA outputs are separately normalized and then blended to obtain a continuous two-dimensional position for each sensor; datasets without additional attributes use geographic coordinates alone. These positions guide a one-to-one assignment of sensors to fixed, uniformly spaced points on a square grid in $\Omega = [0, 1]^2$. The assigned grid points serve as sensor coordinates for both the encoder and the query decoder. The assignment grid and sensor allocation are reconstructed for each period's sensor set, while the latent field grid and learned parameter shapes remain fixed.

Optimization. We use AdamW with learning rates of 5×10^{-3} for PEMS-Stream and 10^{-2} for CA-Stream and AIR-Stream, a batch size of 64, and gradient norm clipping at 5.0. Training runs for at most 200 epochs per period and stops early if validation MAE fails to improve for more than 10 consecutive epochs. The training objective is masked MAE, computed in standardized target space:

$$\mathcal{L}_{\mathrm{MAE}} = \frac{\sum_{i,t} m_{it} \, |\hat{y}_{it} - y_{it}|}{\sum_{i,t} m_{it} + \varepsilon}, \tag{49}$$

where m_{it} indicates finite target entries after preprocessing, and the sums run over the mini-batch.

Continual protocol. We train the model from scratch in period 1. For each subsequent period, we initialize it from the selected parameters of the previous period and fine-tune the same shared parameter set on current training samples. Observation histories and coordinate assignments are updated for the current sensor set, while the latent-grid dimensions and learned parameter shapes remain unchanged.

Metrics and evaluation protocol. We evaluate forecasting performance using MAE, RMSE, and MAPE. All experiments follow the stream construction and evaluation protocol of Liu & Zhang (2026), using the same data splits, prediction horizons, and evaluation masks across methods.

E ALGORITHMS

Algorithm 1 summarizes the forward pass described in Section 3, while Algorithm 2 presents the continual training protocol described in Sections 3.5 and 4.1. Both algorithms follow the notation in Appendix B and use the optimization settings specified in Appendix D.

F BASELINE DESCRIPTIONS

Section 4.1 compares **STFO** with four families of methods. We describe each baseline and its role in the comparison. All methods follow the evaluation protocol, prediction horizons, data splits, and evaluation masks specified in Appendix D.

❶ ***General spatio-temporal forecasting.*** GWNET (Wu et al., 2019) combines graph convolution with dilated causal temporal convolution and learns a self-adaptive adjacency matrix to capture spatial dependencies without a predefined graph. STID (Shao et al., 2022) incorporates spatial and

Algorithm 1 **STFO** forward pass for one period

Input: histories $\{\boldsymbol{u}_i\}_{i\in\mathcal{V}_p}$, coordinates $\{\boldsymbol{x}_i\}_{i\in\mathcal{V}_p}$, query set $\mathcal{V}_q \subseteq \mathcal{V}_p$, parameters θ
Output: forecasts $\{\hat{\boldsymbol{y}}_q\}_{q\in\mathcal{V}_q}$

$\boldsymbol{e}_i \leftarrow \tau_\theta(\boldsymbol{u}_i)$ for all $i \in \mathcal{V}_p$ ▷ temporal encoding
for $g = 1$ to G **do**
  $a_{gi} \leftarrow \kappa_{\sigma_{\mathrm{enc}}}(\boldsymbol{\xi}_g, \boldsymbol{x}_i) \big/ \sum_{j\in\mathcal{V}_p} \kappa_{\sigma_{\mathrm{enc}}}(\boldsymbol{\xi}_g, \boldsymbol{x}_j)$
  $\boldsymbol{r}_g \leftarrow \sum_{i\in\mathcal{V}_p} a_{gi}\boldsymbol{e}_i$ ▷ normalized kernel lift
  $\boldsymbol{z}_g \leftarrow \psi_\theta([\boldsymbol{r}_g \,\|\, \rho(\boldsymbol{\xi}_g)])$ ▷ position-aware latent code
end for
$\bar{z}_g \leftarrow d^{-1}\mathbf{1}_d^\top \boldsymbol{z}_g$; $\widehat{\bar{z}} \leftarrow \mathrm{DFT}(\bar{z})$
$E_m \leftarrow |\widehat{\bar{z}}(\boldsymbol{\omega}_m)|^2 + \epsilon_s$ and $\omega_m \leftarrow \|\boldsymbol{\omega}_m\|_2 + \epsilon_f$ for all $m \in \mathcal{J}$
$\boldsymbol{c} \leftarrow [\alpha, \tilde{\beta}, \eta_1, \ldots, \eta_B]^\top$ ▷ regime descriptor, Eq. (3)
for $\ell = 1$ to L **do**
  $(\boldsymbol{\gamma}, \boldsymbol{\delta}) \leftarrow 0.1\tanh(\mathrm{MLP}_{\theta_\ell}(\boldsymbol{c}))$; $\mathbf{Z}^{\mathrm{c}} \leftarrow \mathbf{Z} \odot (\mathbf{1}_G(\mathbf{1}_d + \boldsymbol{\gamma})^\top) + \mathbf{1}_G\boldsymbol{\delta}^\top$ ▷ FiLM conditioning
  $\mathbf{Z} \leftarrow \mathbf{Z} + \mathcal{P}_{\theta_\ell}(\mathbf{Z}^{\mathrm{c}}) + \mathcal{A}_{\theta_\ell}(\mathbf{Z}^{\mathrm{c}})$ ▷ spectral and attention paths
  $\mathbf{Z} \leftarrow \mathrm{LN}_\ell(\mathbf{Z}) + \mathbf{1}_G\boldsymbol{b}_{\mathrm{res},\ell}^\top$ ▷ residual normalization
end for
$\mathbf{H} \leftarrow \mathbf{Z}$
for $q \in \mathcal{V}_q$ **do**
  $b_{qg} \leftarrow \kappa_{\sigma_{\mathrm{dec}}}(\boldsymbol{x}_q, \boldsymbol{\xi}_g) \big/ \sum_r \kappa_{\sigma_{\mathrm{dec}}}(\boldsymbol{x}_q, \boldsymbol{\xi}_r)$
  $\tilde{\boldsymbol{h}}_q \leftarrow \sum_g b_{qg}\boldsymbol{h}_g$
  $\hat{\boldsymbol{y}}_q \leftarrow \mathbf{W}_{\mathrm{out}}\,\mathrm{LN}(\tilde{\boldsymbol{h}}_q) + \boldsymbol{b}_{\mathrm{out}} + \mathbf{W}_{\mathrm{skip}}\boldsymbol{u}_q + \boldsymbol{b}_{\mathrm{skip}}$ ▷ coordinate query
end for
return $\{\hat{\boldsymbol{y}}_q\}_{q\in\mathcal{V}_q}$

temporal identity embeddings into a lightweight predictor, demonstrating that a simple architecture with explicit identity information can remain competitive with more elaborate designs. ITRANSFORMER (Liu et al., 2023c) inverts the attention axis of a transformer by treating each series as a token, enabling effective multivariate forecasting without an explicit spatial prior. Together, these methods represent graph-based, identity-based, and sequence-based approaches to forecasting from the same observations.

❷ ***Neural operators.*** DEEPONET (Lu et al., 2021) approximates nonlinear operators through a branch–trunk factorization: the branch encodes the input function, while the trunk encodes the query coordinate. This architecture is supported by a universal approximation theorem for operators. FNO (Li et al., 2020) parameterizes a global convolution kernel in the spectral domain and retains a small set of Fourier modes, allowing transfer across resolutions. Both methods share the operator perspective of **STFO**, mapping input functions to output functions. In our experiments, they are trained on fixed meshes and serve as operator baselines for comparison with forecasting on irregular, expanding sensor layouts.

❸ ***Continual spatio-temporal learning.*** TRAFFICSTREAM (Chen et al., 2021) combines experience replay with parameter smoothing for growing road networks. STKEC (Wang et al., 2023a) expands and consolidates task-specific knowledge, enabling new sensors to reuse learned structure. PECPM (Wang et al., 2023b) expands and consolidates a bank of spatio-temporal patterns as the graph evolves. STRAP (Zhang et al., 2025a) retrieves spatio-temporal patterns to improve out-of-distribution generalization. EAC (Chen & Liang, 2025a) studies continual tuning by expanding and compressing adapted parameters. EVOLVE (Yang et al., 2027) addresses changing traffic-network topologies through continual learning. All six retain the current sensor set in their representations; **STFO** treats expansion as changing sampling locations on a fixed latent field.

❹ ***Online adaptation.*** DOL (Wang et al., 2025) uses AST-Net, which augments a frozen Graph WaveNet backbone with location-specific learners that adapt online. We implement ST-TTC (Chen & Liang, 2025b) on an EAC backbone and calibrate its forecasts through online updates to spectral amplitude and phase offsets.

Algorithm 2 Continual fine-tuning of **STFO** across the sensor stream

Input: period data $\{(\mathcal{V}_p, \boldsymbol{s}_{\mathcal{V}_p}, \mathcal{D}_p)\}_{p=1}^{P}$, learning rate η, batch size B_s, maximum epochs $E_{\max}$, patience E_{pat}
Output: parameters $\theta^{(P)}$

initialize θ from scratch
for $p = 1$ to P **do**
  $\boldsymbol{x}_i^{(p)} \leftarrow \chi_p(\boldsymbol{s}_i)$ for all $i \in \mathcal{V}_p$
  if $p > 1$ **then**
    $\theta \leftarrow \theta^{(p-1)}$
  end if
  reset AdamW state; initialize checkpoint set $\mathcal{H}_p \leftarrow \varnothing$
  for epoch $= 1$ to $E_{\max}$ **do**
    for all mini-batches $\mathcal{B}$ from the training split of $\mathcal{D}_p$ with batch size B_s **do**
      $\{\hat{\boldsymbol{y}}_q\} \leftarrow \textsc{Forward}(\mathcal{B}; \{\boldsymbol{x}_i^{(p)}\}_{i \in \mathcal{V}_p}, \theta)$ ▷ Algorithm 1
      $\mathcal{L} \leftarrow \sum_{i,t} m_{it} |\hat{y}_{it} - y_{it}| / (\sum_{i,t} m_{it} + \varepsilon)$ ▷ masked MAE, Eq. (49)
      update θ using AdamW with learning rate η and gradient clipping
    end for
    evaluate MAE on the validation split of $\mathcal{D}_p$; save θ in $\mathcal{H}_p$ if it improves the best value so far
    if validation MAE has not improved for E_{pat} epochs **then**
      break
    end if
  end for
  select $\theta^{(p)}$ from $\mathcal{H}_p$ by validation MAE with optional checkpoint parameter averaging
end for
return $\theta^{(P)}$

G Statistical Significance Analysis

We compare **STFO**-Large with 13 baselines using one-sided Wilcoxon signed-rank tests on 27 paired entries per metric, with Holm correction across baselines within each metric. Table 4 reports adjusted p-values below 0.0001 for all comparisons except MAPE against DOL ($p_{\text{Holm}} < 0.01$). We report these p-values for reference (some paired entries are dependent).

H Full New-Sensor Adaptation Results

We evaluate performance on existing and newly introduced sensors across expansion periods. Zero-shot evaluation uses the model trained in the preceding period without adaptation, whereas full adaptation fine-tunes that model using all available current-period supervision.

PEMS-Stream. Tables 5, 6, and 7 show that **STFO**-Large leads across all metrics, horizons, sensor groups, and six periods under full adaptation. Its zero-shot advantage is stronger on existing sensors, with the lowest average MAE and RMSE in five periods. On new sensors, it achieves the lowest average RMSE in 2012, 2015, and 2017, and the lowest average MAE and MAPE in 2015.

CA-Stream. Tables 8, 9, and 10 show that **STFO**-Large ranks first across all three periods, metrics, horizons, and sensor groups under full adaptation. Under zero-shot evaluation, it achieves the lowest average MAE, RMSE, and MAPE on existing sensors in P1 and P2. On new sensors, it leads in average MAPE in P1 and in all three average metrics in P2.

AIR-Stream. Tables 11 and 12 complement Table 2. In zero-shot evaluation, **STFO**-Large leads in average RMSE and MAPE on existing sensors in every period and on new sensors in 2017 and 2018. With full adaptation, it leads in RMSE and MAPE at every horizon for both groups in 2017 and 2018; in 2019, it leads in average MAPE for both groups but not in RMSE at every horizon.

Overall, **STFO**-Large achieves the lowest errors across all periods, horizons, and sensor groups on both traffic streams under full adaptation.

Table 4: Wilcoxon signed-rank test results (one-sided, H_1: STFO-L $<$ Baseline) for the main forecasting results. $N = 27$ paired entries per comparison (dataset $\times$ horizon $\times$ seed). Shading and stars indicate nominal Holm-adjusted levels.

Baseline	N	**STFO-L Wins**	**Baseline Wins**	**Ties**	**Mean Δ%**	W^+	p**-value**	p_{Holm}
			Metric: MAE	H_1: STFO-L $<$ Baseline				
GWNet	27	27	0	0	37.07%	0.0	< 0.0001	$< 0.0001^{***}$
STID	27	27	0	0	48.23%	0.0	< 0.0001	$< 0.0001^{***}$
iTrans	27	27	0	0	36.40%	0.0	< 0.0001	$< 0.0001^{***}$
DeepONet	27	27	0	0	11.84%	0.0	< 0.0001	$< 0.0001^{***}$
FNO	27	26	1	0	13.25%	1.0	< 0.0001	$< 0.0001^{***}$
TrafficStream	27	27	0	0	27.06%	0.0	< 0.0001	$< 0.0001^{***}$
STKEC	27	27	0	0	27.07%	0.0	< 0.0001	$< 0.0001^{***}$
PECPM	27	27	0	0	26.89%	0.0	< 0.0001	$< 0.0001^{***}$
STRAP	27	27	0	0	17.87%	0.0	< 0.0001	$< 0.0001^{***}$
EAC	27	27	0	0	24.00%	0.0	< 0.0001	$< 0.0001^{***}$
EVOLVE	27	27	0	0	18.18%	0.0	< 0.0001	$< 0.0001^{***}$
DOL	27	24	3	0	4.27%	18.0	< 0.0001	$< 0.0001^{***}$
ST-TTC	27	27	0	0	14.79%	0.0	< 0.0001	$< 0.0001^{***}$
			Metric: RMSE	H_1: STFO-L $<$ Baseline				
GWNet	27	27	0	0	33.28%	0.0	< 0.0001	$< 0.0001^{***}$
STID	27	27	0	0	44.32%	0.0	< 0.0001	$< 0.0001^{***}$
iTrans	27	27	0	0	32.68%	0.0	< 0.0001	$< 0.0001^{***}$
DeepONet	27	27	0	0	10.67%	0.0	< 0.0001	$< 0.0001^{***}$
FNO	27	26	1	0	12.15%	1.0	< 0.0001	$< 0.0001^{***}$
TrafficStream	27	27	0	0	23.58%	0.0	< 0.0001	$< 0.0001^{***}$
STKEC	27	27	0	0	23.84%	0.0	< 0.0001	$< 0.0001^{***}$
PECPM	27	27	0	0	23.43%	0.0	< 0.0001	$< 0.0001^{***}$
STRAP	27	27	0	0	17.27%	0.0	< 0.0001	$< 0.0001^{***}$
EAC	27	27	0	0	19.80%	0.0	< 0.0001	$< 0.0001^{***}$
EVOLVE	27	27	0	0	17.35%	0.0	< 0.0001	$< 0.0001^{***}$
DOL	27	27	0	0	4.93%	0.0	< 0.0001	$< 0.0001^{***}$
ST-TTC	27	27	0	0	11.79%	0.0	< 0.0001	$< 0.0001^{***}$
			Metric: MAPE	H_1: STFO-L $<$ Baseline				
GWNet	27	27	0	0	36.29%	0.0	< 0.0001	$< 0.0001^{***}$
STID	27	27	0	0	49.18%	0.0	< 0.0001	$< 0.0001^{***}$
iTrans	27	27	0	0	40.43%	0.0	< 0.0001	$< 0.0001^{***}$
DeepONet	27	26	1	0	17.11%	2.0	< 0.0001	$< 0.0001^{***}$
FNO	27	24	3	0	18.16%	15.0	< 0.0001	$< 0.0001^{***}$
TrafficStream	27	27	0	0	20.97%	0.0	< 0.0001	$< 0.0001^{***}$
STKEC	27	27	0	0	20.13%	0.0	< 0.0001	$< 0.0001^{***}$
PECPM	27	27	0	0	21.22%	0.0	< 0.0001	$< 0.0001^{***}$
STRAP	27	27	0	0	15.01%	0.0	< 0.0001	$< 0.0001^{***}$
EAC	27	27	0	0	19.80%	0.0	< 0.0001	$< 0.0001^{***}$
EVOLVE	27	26	1	0	16.01%	1.0	< 0.0001	$< 0.0001^{***}$
DOL	27	20	7	0	2.87%	80.0	0.0038	0.0038^{**}
ST-TTC	27	27	0	0	20.39%	0.0	< 0.0001	$< 0.0001^{***}$

$^{***}p_{\mathrm{Holm}} < 0.001$ $^{**}p_{\mathrm{Holm}} < 0.01$. Holm correction is applied across 13 baselines per metric.

Mean Δ% averages relative error reductions across nine dataset–horizon settings using three-seed means, excluding `Avg.` rows. W^+ is the positive-rank sum for STFO-L $-$ Baseline.

Table 5: Period-wise MAE on existing and newly introduced sensors of PEMS-Stream. Results are mean ± standard deviation over three runs. Lower is better; **bold** and underlined denote the best and second-best methods within each period, sensor group, adaptation budget, and horizon.

Budget	Sensor group	FNO				EVOLVE				STFO-Large			
		$h=3$	$h=6$	$h=12$	*Avg.*	$h=3$	$h=6$	$h=12$	*Avg.*	$h=3$	$h=6$	$h=12$	*Avg.*
2012: 655 existing sensors; 60 newly introduced sensors													
Zero-shot	Existing	**12.46**±0.08	**14.14**±0.04	17.77±0.04	14.44±0.04	12.97±0.13	14.64±0.24	19.12±0.35	15.12±0.24	**12.46**±0.29	14.25±0.48	**17.28**±0.65	**14.34**±0.43
	Newly introduced	13.44±0.13	15.93±0.14	21.44±0.35	16.40±0.16	**12.74**±0.13	**14.39**±0.27	**18.46**±0.41	**14.76**±0.26	12.85±0.34	15.08±0.55	18.79±0.97	15.19±0.57
Full	Existing	12.01±0.04	13.56±0.04	16.98±0.08	13.84±0.04	12.88±0.09	14.57±0.17	19.04±0.30	15.05±0.18	**10.39**±0.05	**10.90**±0.04	**11.96**±0.13	**10.94**±0.05
	Newly introduced	12.47±0.03	14.55±0.04	19.24±0.05	14.96±0.04	12.57±0.08	14.17±0.14	18.09±0.30	14.52±0.15	**10.24**±0.05	**10.83**±0.04	**11.97**±0.14	**10.86**±0.06
2013: 715 existing sensors; 71 newly introduced sensors													
Zero-shot	Existing	**12.96**±0.05	**14.97**±0.07	**19.38**±0.06	**15.33**±0.04	13.43±0.12	15.42±0.18	20.42±0.30	15.92±0.19	13.57±0.21	16.04±0.23	20.38±0.53	16.19±0.27
	Newly introduced	12.03±0.23	14.04±0.13	19.94±0.28	14.81±0.19	**10.89**±0.04	**12.20**±0.18	**16.08**±0.38	**12.67**±0.19	11.23±0.12	13.36±0.12	16.85±0.36	13.43±0.09
Full	Existing	12.37±0.04	14.17±0.03	18.10±0.02	14.51±0.03	13.14±0.12	15.01±0.13	19.67±0.16	15.47±0.12	**10.47**±0.03	**11.00**±0.05	**11.99**±0.12	**11.02**±0.06
	Newly introduced	10.59±0.03	12.26±0.02	16.44±0.05	12.71±0.01	10.48±0.12	11.44±0.14	14.56±0.22	11.85±0.15	**8.27**±0.06	**8.63**±0.07	**9.34**±0.13	**8.65**±0.08
2014: 786 existing sensors; 36 newly introduced sensors													
Zero-shot	Existing	**14.47**±0.16	16.20±0.10	20.38±0.23	16.63±0.14	14.78±0.15	16.51±0.08	21.17±0.14	17.01±0.13	14.54±0.29	**16.08**±0.27	**18.96**±0.08	**16.16**±0.19
	Newly introduced	12.75±0.23	14.69±0.19	19.63±0.35	15.26±0.24	**11.42**±0.35	**12.40**±0.22	**15.30**±0.25	**12.75**±0.29	12.56±0.74	13.83±0.84	16.04±0.48	13.84±0.66
Full	Existing	13.14±0.07	14.81±0.01	18.86±0.02	15.21±0.01	13.77±0.13	15.61±0.13	20.38±0.07	16.10±0.11	**10.91**±0.03	**11.29**±0.02	**12.18**±0.02	**11.34**±0.01
	Newly introduced	10.52±0.08	12.20±0.14	16.88±0.22	12.80±0.14	9.63±0.26	10.70±0.26	13.60±0.24	11.02±0.25	**7.77**±0.02	**8.05**±0.01	**8.73**±0.02	**8.09**±0.00
2015: 822 existing sensors; 12 newly introduced sensors													
Zero-shot	Existing	13.21±0.06	14.88±0.02	18.85±0.09	15.24±0.01	13.53±0.20	15.40±0.21	19.70±0.14	15.76±0.18	**12.09**±0.11	**13.19**±0.11	**15.07**±0.09	**13.21**±0.08
	Newly introduced	8.60±0.04	9.61±0.08	12.28±0.04	9.91±0.03	8.36±0.09	9.33±0.10	12.08±0.33	9.72±0.12	**7.94**±0.05	**8.72**±0.07	**10.28**±0.10	**8.83**±0.07
Full	Existing	12.79±0.04	14.60±0.03	18.42±0.02	14.89±0.03	13.39±0.07	15.23±0.08	19.61±0.12	15.64±0.09	**10.86**±0.01	**11.38**±0.04	**12.33**±0.07	**11.39**±0.03
	Newly introduced	8.00±0.02	8.86±0.01	10.95±0.05	9.09±0.02	8.37±0.13	9.27±0.15	12.03±0.27	9.69±0.19	**7.13**±0.03	**7.40**±0.04	**8.08**±0.06	**7.48**±0.03
2016: 834 existing sensors; 16 newly introduced sensors													
Zero-shot	Existing	12.53±0.02	14.39±0.05	18.56±0.10	14.73±0.04	12.88±0.05	14.62±0.02	18.97±0.08	15.06±0.01	**12.52**±0.16	**14.37**±0.29	**17.54**±0.27	**14.47**±0.14
	Newly introduced	13.16±0.17	15.83±0.05	22.19±0.48	16.44±0.22	**11.48**±0.07	**12.99**±0.04	**16.31**±0.03	**13.28**±0.03	11.91±0.39	13.87±0.30	17.09±0.54	13.92±0.30
Full	Existing	12.16±0.03	13.89±0.04	17.69±0.03	14.20±0.02	12.72±0.05	14.43±0.05	18.60±0.05	14.83±0.02	**10.31**±0.01	**10.81**±0.01	**11.80**±0.03	**10.84**±0.01
	Newly introduced	12.32±0.05	14.84±0.08	20.51±0.12	15.34±0.01	11.22±0.02	12.63±0.05	15.58±0.13	12.86±0.05	**9.54**±0.03	**9.94**±0.04	**10.83**±0.07	**9.98**±0.03
2017: 850 existing sensors; 21 newly introduced sensors													
Zero-shot	Existing	20.27±0.29	21.66±0.17	25.76±0.15	22.21±0.10	**19.28**±0.84	20.71±0.69	24.40±0.53	21.09±0.66	19.61±0.49	**20.69**±0.49	**22.72**±0.75	**20.80**±0.43
	Newly introduced	14.90±0.35	14.94±0.63	16.37±0.23	15.37±0.18	**13.77**±0.71	**14.19**±0.62	15.91±0.44	**14.45**±0.61	14.36±0.41	14.52±0.40	**15.06**±0.35	14.65±0.33
Full	Existing	13.72±0.10	15.64±0.07	19.98±0.11	16.02±0.08	14.07±0.05	16.05±0.03	20.63±0.09	16.44±0.04	**11.80**±0.02	**12.42**±0.04	**13.61**±0.07	**12.46**±0.04
	Newly introduced	8.69±0.08	9.94±0.12	13.19±0.17	10.31±0.11	8.66±0.04	9.65±0.07	12.53±0.18	10.01±0.07	**7.52**±0.03	**8.06**±0.10	**9.17**±0.09	**8.15**±0.07

Table 6: Period-wise RMSE on existing and newly introduced sensors of PEMS-Stream. Results are mean ± standard deviation over three runs. Lower is better; **bold** and underlined denote the best and second-best methods within each period, sensor group, adaptation budget, and horizon.

Budget	Sensor group	FNO				EVOLVE				STFO-Large			
		$h=3$	$h=6$	$h=12$	*Avg.*	$h=3$	$h=6$	$h=12$	*Avg.*	$h=3$	$h=6$	$h=12$	*Avg.*
2012: 655 existing sensors; 60 newly introduced sensors													
Zero-shot	Existing	19.35±0.05	22.25±0.01	28.38±0.04	22.73±0.02	20.56±0.32	23.35±0.51	30.36±0.59	24.04±0.48	**19.25**±0.31	**22.02**±0.55	**26.95**±0.69	**22.21**±0.48
	Newly introduced	21.37±0.08	25.31±0.20	33.42±0.51	25.89±0.20	21.54±0.30	24.32±0.52	31.08±0.64	24.91±0.49	**20.86**±0.24	**24.27**±0.49	**30.40**±0.87	**24.52**±0.52
Full	Existing	18.97±0.07	21.73±0.05	27.62±0.10	22.18±0.06	20.27±0.18	23.01±0.29	29.96±0.40	23.71±0.28	**16.33**±0.07	**17.16**±0.08	**18.77**±0.21	**17.19**±0.09
	Newly introduced	20.69±0.08	24.16±0.07	31.29±0.09	24.64±0.07	21.08±0.14	23.74±0.30	30.39±0.47	24.36±0.28	**17.38**±0.07	**18.41**±0.09	**20.37**±0.27	**18.43**±0.11
2013: 715 existing sensors; 71 newly introduced sensors													
Zero-shot	Existing	**20.90**±0.09	**24.55**±0.09	**31.76**±0.13	**24.98**±0.05	21.66±0.19	25.00±0.26	32.53±0.34	25.59±0.25	22.08±0.42	26.35±0.57	33.81±1.09	26.60±0.64
	Newly introduced	19.02±0.19	22.53±0.13	31.43±0.29	23.43±0.19	**18.58**±0.21	**21.04**±0.40	**27.80**±0.58	**21.78**±0.37	18.98±0.25	22.68±0.26	29.52±0.04	23.03±0.19
Full	Existing	20.12±0.05	23.49±0.06	30.09±0.03	23.92±0.05	21.33±0.20	24.42±0.29	31.60±0.36	25.03±0.26	**17.14**±0.03	**18.23**±0.07	**19.79**±0.15	**18.15**±0.07
	Newly introduced	17.91±0.05	21.05±0.05	28.84±0.08	21.84±0.05	17.79±0.16	19.42±0.30	24.95±0.56	20.15±0.30	**13.74**±0.07	**14.36**±0.11	**15.44**±0.18	**14.35**±0.11
2014: 786 existing sensors; 36 newly introduced sensors													
Zero-shot	Existing	22.14±0.14	25.26±0.14	32.47±0.22	25.90±0.14	23.12±0.18	26.19±0.29	33.73±0.35	26.89±0.25	**22.02**±0.23	**24.78**±0.16	**29.82**±0.24	**24.94**±0.16
	Newly introduced	18.92±0.14	22.09±0.19	30.58±0.28	23.10±0.20	**17.73**±0.15	**19.75**±0.18	25.28±0.30	**20.33**±0.16	18.41±0.60	20.84±0.70	**25.20**±0.76	20.97±0.62
Full	Existing	21.24±0.02	24.39±0.06	31.42±0.07	24.95±0.05	22.07±0.27	25.08±0.33	32.56±0.37	25.78±0.29	**17.62**±0.03	**18.30**±0.03	**19.63**±0.06	**18.31**±0.03
	Newly introduced	17.24±0.10	20.16±0.13	27.88±0.23	21.07±0.15	16.78±0.31	18.82±0.44	24.28±0.63	19.37±0.42	**13.29**±0.04	**13.75**±0.05	**14.84**±0.07	**13.82**±0.04
2015: 822 existing sensors; 12 newly introduced sensors													
Zero-shot	Existing	21.61±0.04	25.02±0.08	31.90±0.17	25.45±0.02	22.13±0.34	25.45±0.34	32.38±0.26	25.92±0.31	**20.11**±0.14	**22.24**±0.16	**25.60**±0.20	**22.21**±0.11
	Newly introduced	12.43±0.04	14.10±0.06	18.20±0.07	14.52±0.04	12.14±0.20	13.80±0.32	18.49±0.73	14.42±0.36	**11.60**±0.12	**12.82**±0.20	**15.26**±0.24	**12.99**±0.17
Full	Existing	21.01±0.05	24.55±0.10	31.25±0.06	24.92±0.07	21.66±0.04	24.92±0.09	32.01±0.17	25.46±0.10	**17.76**±0.02	**18.71**±0.03	**20.22**±0.08	**18.64**±0.04
	Newly introduced	11.57±0.03	12.94±0.07	16.22±0.09	13.29±0.05	12.12±0.18	13.72±0.25	18.59±0.61	14.42±0.34	**10.43**±0.03	**10.85**±0.03	**11.80**±0.02	**10.94**±0.03
2016: 834 existing sensors; 16 newly introduced sensors													
Zero-shot	Existing	22.79±0.02	26.35±0.08	33.54±0.09	26.78±0.03	23.06±0.05	26.19±0.06	32.83±0.08	26.65±0.04	**22.75**±0.26	**26.06**±0.51	**31.32**±0.25	**26.09**±0.26
	Newly introduced	22.85±0.11	26.95±0.18	36.87±0.52	27.83±0.13	**21.80**±0.21	**25.39**±0.14	**31.39**±0.10	**25.46**±0.14	22.83±0.55	26.50±0.52	32.57±0.76	26.47±0.32
Full	Existing	22.47±0.02	25.85±0.02	32.59±0.10	26.25±0.04	22.87±0.14	25.95±0.10	32.44±0.11	26.35±0.06	**19.61**±0.07	**20.82**±0.05	**22.84**±0.07	**20.76**±0.05
	Newly introduced	22.51±0.11	26.60±0.32	35.29±0.31	27.09±0.22	21.69±0.07	24.94±0.08	30.30±0.22	24.91±0.09	**18.77**±0.20	**19.60**±0.23	**21.66**±0.18	**19.57**±0.18
2017: 850 existing sensors; 21 newly introduced sensors													
Zero-shot	Existing	28.47±0.33	31.54±0.15	38.83±0.29	32.22±0.20	**27.45**±0.63	**30.23**±0.49	36.69±0.52	30.78±0.45	27.84±0.37	30.34±0.46	**34.56**±1.02	**30.45**±0.46
	Newly introduced	18.99±0.33	19.96±0.54	23.28±0.32	20.48±0.20	18.50±0.59	19.85±0.50	24.37±0.70	20.48±0.52	**18.33**±0.55	**19.26**±0.63	**21.35**±0.78	**19.46**±0.54
Full	Existing	22.78±0.09	26.09±0.07	32.81±0.12	26.55±0.06	23.39±0.11	26.70±0.09	33.64±0.18	27.15±0.10	**19.92**±0.02	**21.06**±0.06	**22.99**±0.08	**21.05**±0.05
	Newly introduced	13.41±0.11	15.18±0.18	19.61±0.21	15.64±0.15	13.93±0.07	15.88±0.09	20.93±0.25	16.43±0.11	**11.98**±0.05	**12.89**±0.15	**14.80**±0.21	**13.03**±0.15

Table 7: Period-wise MAPE on existing and newly introduced sensors of PEMS-Stream. Results are mean ± standard deviation over three runs. Lower is better; **bold** and underlined denote the best and second-best methods within each period, sensor group, adaptation budget, and horizon.

Budget	Sensor group	FNO				EVOLVE				STFO-Large			
		$h=3$	$h=6$	$h=12$	*Avg.*	$h=3$	$h=6$	$h=12$	*Avg.*	$h=3$	$h=6$	$h=12$	*Avg.*
2012: 655 existing sensors; 60 newly introduced sensors													
Zero-shot	Existing	21.59±0.58	26.20±0.58	34.47±0.46	26.49±0.26	**17.41**±0.60	**19.82**±1.46	**26.06**±2.17	**20.61**±1.33	18.51±1.57	21.93±2.27	26.53±3.35	21.86±2.13
	Newly introduced	37.93±1.59	50.17±2.23	75.31±4.35	51.68±2.21	**22.66**±0.69	**25.49**±1.83	**32.22**±2.69	**26.30**±1.81	29.49±4.01	37.09±5.86	46.44±9.29	36.83±5.87
Full	Existing	18.29±0.27	22.31±0.33	30.56±0.64	23.05±0.37	17.55±0.62	19.66±0.66	25.33±2.01	20.35±0.91	**13.83**±0.12	**15.11**±0.22	**16.33**±0.27	**14.92**±0.15
	Newly introduced	27.34±0.14	34.94±0.65	53.48±0.36	36.95±0.37	23.39±1.75	25.51±0.90	31.02±1.16	26.15±0.41	**18.14**±0.14	**19.89**±0.34	**21.58**±0.43	**19.64**±0.25
2013: 715 existing sensors; 71 newly introduced sensors													
Zero-shot	Existing	19.10±0.44	22.82±0.41	31.72±0.62	23.81±0.47	17.35±0.72	**19.56**±0.92	25.47±2.21	20.29±1.20	**16.56**±0.47	19.90±1.17	**23.60**±1.37	**19.54**±0.97
	Newly introduced	46.68±3.65	57.86±2.53	94.61±4.79	63.13±3.55	**26.62**±2.26	**29.24**±1.67	**36.94**±4.94	**30.40**±2.35	27.83±1.60	36.05±3.52	42.57±4.31	34.37±3.13
Full	Existing	18.97±0.12	22.82±0.40	32.85±0.54	24.05±0.27	16.46±0.44	18.27±0.33	23.32±0.12	18.92±0.31	**13.57**±0.25	**14.28**±0.30	**15.50**±0.27	**14.20**±0.14
	Newly introduced	27.51±0.17	33.74±0.77	48.44±1.76	35.37±0.75	25.48±1.68	26.98±1.29	32.25±0.81	27.86±1.31	**19.82**±0.48	**20.72**±0.46	**22.33**±0.55	**20.53**±0.33
2014: 786 existing sensors; 36 newly introduced sensors													
Zero-shot	Existing	36.03±0.94	39.21±0.95	47.12±2.02	40.36±1.02	**32.44**±2.94	**34.04**±2.49	39.14±3.25	**34.61**±2.96	35.39±1.30	36.63±1.98	**38.34**±0.51	36.29±1.18
	Newly introduced	112.12±3.20	128.48±3.32	172.73±8.48	135.14±4.67	**83.20**±10.56	**84.70**±8.02	**92.53**±10.99	**85.65**±10.26	102.67±7.20	107.38±10.57	112.76±5.44	106.12±7.63
Full	Existing	21.88±1.66	24.17±0.50	35.04±0.97	26.12±0.29	18.41±1.72	20.26±1.47	25.25±0.75	20.85±1.27	**14.66**±0.04	**15.25**±0.15	**16.55**±0.20	**15.26**±0.07
	Newly introduced	63.52±3.16	76.98±2.67	126.24±3.67	85.03±2.10	33.84±7.49	35.27±6.04	39.96±1.79	35.96±5.27	**26.01**±0.04	**27.75**±0.68	**31.28**±1.61	**27.73**±0.49
2015: 822 existing sensors; 12 newly introduced sensors													
Zero-shot	Existing	22.86±1.82	25.17±0.59	36.50±0.78	27.17±0.25	18.40±1.87	20.20±1.56	24.86±0.71	20.71±1.36	**15.50**±0.13	**16.88**±0.15	**19.20**±0.21	**16.85**±0.06
	Newly introduced	28.86±2.69	29.39±0.82	39.70±1.56	31.79±0.20	25.03±2.96	26.88±2.37	33.38±1.11	28.02±1.98	**23.43**±0.50	**26.21**±0.98	**30.32**±1.94	**26.13**±0.95
Full	Existing	21.07±0.34	25.37±0.75	36.43±0.21	26.46±0.22	18.33±0.58	19.93±0.52	25.41±0.84	20.72±0.57	**14.98**±0.32	**15.97**±0.41	**17.52**±0.46	**15.87**±0.21
	Newly introduced	25.93±0.66	29.27±1.37	37.65±1.10	29.95±0.66	25.24±1.25	26.59±1.32	33.42±2.40	27.92±1.47	**21.67**±0.52	**23.07**±0.87	**25.71**±1.16	**23.14**±0.51
2016: 834 existing sensors; 16 newly introduced sensors													
Zero-shot	Existing	19.13±0.41	23.05±0.21	32.83±0.59	23.92±0.20	16.71±0.45	18.19±0.34	23.54±0.72	19.00±0.49	**15.19**±0.55	**17.10**±0.18	**20.55**±0.17	**17.17**±0.18
	Newly introduced	86.89±2.99	118.45±0.89	203.33±9.53	128.20±4.22	**33.11**±4.09	**34.88**±2.84	**44.41**±3.20	**36.69**±3.03	34.35±1.73	40.19±1.30	49.76±0.92	40.26±1.08
Full	Existing	18.37±0.14	23.31±1.00	33.54±1.23	24.03±0.47	16.10±0.24	17.90±0.31	22.75±0.54	18.48±0.30	**13.27**±0.23	**13.98**±0.35	**15.51**±0.25	**14.03**±0.27
	Newly introduced	61.47±1.16	89.99±3.22	160.10±6.77	97.83±2.78	30.37±0.29	33.01±0.25	40.49±1.40	34.10±0.35	**28.67**±0.18	**30.06**±0.64	**34.03**±0.25	**30.36**±0.36
2017: 850 existing sensors; 21 newly introduced sensors													
Zero-shot	Existing	45.14±1.47	43.53±0.83	48.77±1.13	45.76±0.71	**32.55**±1.92	**32.38**±2.08	**32.92**±1.71	**32.48**±2.07	41.38±0.64	40.54±0.41	38.21±0.84	40.51±0.12
	Newly introduced	130.38±4.81	119.97±7.48	122.47±1.66	125.81±1.32	**75.49**±7.61	**71.56**±7.31	**64.48**±4.59	**70.71**±7.04	121.79±0.70	116.84±3.65	104.50±4.98	116.58±3.05
Full	Existing	20.18±0.43	24.15±0.42	35.79±0.41	25.53±0.20	17.24±0.22	19.06±0.23	23.58±0.66	19.47±0.36	**14.52**±0.31	**15.40**±0.14	**17.23**±0.58	**15.59**±0.33
	Newly introduced	49.62±1.49	64.60±0.50	98.78±1.90	68.37±0.86	36.25±1.14	37.94±0.89	45.81±3.17	39.12±1.62	**31.06**±0.79	**33.58**±0.55	**39.80**±2.34	**34.60**±0.95

Table 8: Period-wise MAE on existing and newly introduced sensors of CA-Stream. Results are mean ± standard deviation over three runs. Lower is better; **bold** and underlined denote the best and second-best methods within each period, sensor group, adaptation budget, and horizon.

Budget	Sensor group	FNO				EVOLVE				STFO-Large			
		$h=3$	$h=6$	$h=12$	*Avg.*	$h=3$	$h=6$	$h=12$	*Avg.*	$h=3$	$h=6$	$h=12$	*Avg.*
P1: 480 existing sensors; 211 newly introduced sensors													
Zero-shot	Existing	28.95±1.41	31.25±1.85	36.58±2.55	31.80±1.88	29.81±0.27	30.91±0.47	**34.29**±0.73	31.31±0.43	**26.28**±0.24	**28.52**±0.60	34.95±1.22	**29.58**±0.56
	Newly introduced	24.68±3.25	**24.38**±0.70	**26.28**±1.44	**24.94**±0.65	**24.52**±0.40	25.18±0.50	27.35±0.64	25.52±0.45	25.19±1.02	25.15±0.73	28.49±1.34	26.13±0.92
Full	Existing	17.23±0.05	19.75±0.01	24.42±0.11	20.01±0.03	17.79±0.04	20.41±0.09	26.43±0.15	20.96±0.07	**15.12**±0.05	**16.18**±0.10	**17.97**±0.21	**16.17**±0.11
	Newly introduced	9.37±0.01	11.01±0.13	14.21±0.03	11.24±0.04	9.45±0.02	10.85±0.13	14.10±0.15	11.19±0.11	**8.20**±0.09	**8.67**±0.06	**9.68**±0.20	**8.72**±0.07
P2: 691 existing sensors; 484 newly introduced sensors													
Zero-shot	Existing	20.72±0.28	23.43±0.36	28.70±0.68	23.75±0.39	20.60±0.27	22.03±0.10	26.01±0.14	22.50±0.10	**19.59**±0.67	**21.70**±0.92	**25.52**±0.39	**21.89**±0.39
	Newly introduced	18.43±0.35	21.43±0.44	27.57±1.37	21.90±0.62	17.50±0.22	19.21±0.10	23.47±0.10	19.67±0.14	**16.92**±0.31	**18.95**±0.80	**21.99**±0.24	**18.98**±0.30
Full	Existing	14.75±0.04	16.65±0.06	20.50±0.04	16.91±0.04	15.47±0.25	17.46±0.28	22.42±0.39	17.95±0.30	**13.08**±0.05	**13.97**±0.09	**15.80**±0.18	**14.08**±0.10
	Newly introduced	13.30±0.04	14.49±0.06	16.74±0.02	14.60±0.03	14.47±0.07	16.13±0.09	20.36±0.08	16.55±0.07	**12.27**±0.07	**12.90**±0.10	**14.11**±0.15	**12.94**±0.09
P3: 1,175 existing sensors; 523 newly introduced sensors													
Zero-shot	Existing	**17.62**±0.29	**19.11**±0.02	**22.70**±0.11	**19.46**±0.12	17.80±0.31	19.81±0.34	24.75±0.40	20.33±0.33	18.28±0.51	20.65±0.35	24.75±0.67	20.66±0.29
	Newly introduced	18.46±0.25	20.54±0.26	25.47±0.47	21.00±0.34	**17.47**±0.34	**19.28**±0.35	**23.34**±0.38	**19.66**±0.35	17.63±0.41	19.91±0.42	23.79±0.72	19.96±0.37
Full	Existing	14.48±0.06	16.20±0.02	19.33±0.05	16.36±0.03	15.30±0.03	17.51±0.04	22.62±0.06	17.99±0.03	**13.49**±0.08	**14.61**±0.10	**16.67**±0.18	**14.71**±0.11
	Newly introduced	14.13±0.03	15.52±0.03	18.10±0.06	15.67±0.01	14.97±0.03	16.90±0.05	21.09±0.04	17.28±0.04	**13.33**±0.05	**14.18**±0.07	**15.85**±0.11	**14.30**±0.08

Table 9: Period-wise RMSE on existing and newly introduced sensors of CA-Stream. Results are mean ± standard deviation over three runs. Lower is better; **bold** and underlined denote the best and second-best methods within each period, sensor group, adaptation budget, and horizon.

Budget	Sensor group	FNO				EVOLVE				STFO-Large			
		$h=3$	$h=6$	$h=12$	*Avg.*	$h=3$	$h=6$	$h=12$	*Avg.*	$h=3$	$h=6$	$h=12$	*Avg.*
P1: 480 existing sensors; 211 newly introduced sensors													
Zero-shot	Existing	39.07±1.56	43.59±1.78	52.66±3.27	44.18±2.10	39.51±0.17	42.06±0.36	**48.10**±0.63	42.58±0.36	**36.38**±0.33	**41.35**±0.44	51.78±1.51	**42.41**±0.55
	Newly introduced	30.10±2.86	**31.13**±0.41	35.93±2.39	31.91±0.50	**29.38**±0.30	**31.13**±0.36	**35.41**±0.38	**31.62**±0.27	30.15±1.07	31.62±0.93	36.97±1.86	32.53±1.19
Full	Existing	29.13±0.02	32.85±0.05	39.40±0.18	33.14±0.06	30.38±0.03	34.40±0.16	43.00±0.23	35.05±0.13	**26.58**±0.04	**28.30**±0.10	**30.72**±0.20	**28.18**±0.10
	Newly introduced	17.48±0.05	20.30±0.07	25.22±0.05	20.54±0.03	18.26±0.04	20.97±0.02	26.65±0.06	21.46±0.03	**16.10**±0.05	**17.38**±0.10	**19.14**±0.25	**17.28**±0.13
P2: 691 existing sensors; 484 newly introduced sensors													
Zero-shot	Existing	33.57±0.37	37.52±0.40	45.20±1.00	37.95±0.55	35.61±0.24	37.72±0.23	43.50±0.34	38.37±0.13	**31.70**±0.80	**35.02**±1.15	**41.33**±0.78	**35.39**±0.54
	Newly introduced	27.20±0.34	31.82±0.45	40.69±1.68	32.36±0.70	27.54±0.19	30.52±0.09	37.77±0.18	31.28±0.11	**25.56**±0.41	**29.06**±0.82	**34.54**±0.58	**29.18**±0.33
Full	Existing	26.09±0.07	28.81±0.04	34.32±0.05	29.16±0.02	27.60±0.27	30.85±0.35	38.19±0.51	31.47±0.37	**24.23**±0.07	**25.63**±0.12	**28.32**±0.24	**25.73**±0.13
	Newly introduced	20.41±0.10	22.37±0.06	25.85±0.05	22.45±0.04	22.51±0.14	25.34±0.19	32.14±0.23	25.96±0.16	**18.93**±0.10	**20.05**±0.14	**21.91**±0.23	**20.02**±0.13
P3: 1,175 existing sensors; 523 newly introduced sensors													
Zero-shot	Existing	**26.65**±0.35	**29.24**±0.05	**34.56**±0.23	**29.58**±0.18	27.59±0.23	30.99±0.24	38.69±0.26	31.67±0.22	27.66±0.56	31.52±0.48	38.33±0.76	31.66±0.36
	Newly introduced	25.27±0.37	28.61±0.39	35.76±0.58	29.15±0.46	24.35±0.37	**27.42**±0.35	**33.69**±0.30	**27.88**±0.34	**24.19**±0.46	27.91±0.49	34.41±0.73	28.11±0.45
Full	Existing	24.24±0.07	26.95±0.03	31.71±0.06	27.15±0.05	25.68±0.08	29.17±0.12	36.85±0.19	29.81±0.11	**22.96**±0.10	**24.72**±0.15	**27.68**±0.33	**24.78**±0.18
	Newly introduced	20.50±0.03	22.61±0.03	26.40±0.09	22.81±0.02	22.01±0.05	25.12±0.08	31.33±0.02	25.56±0.06	**19.39**±0.07	**20.72**±0.10	**23.15**±0.17	**20.85**±0.11

Table 10: Period-wise MAPE on existing and newly introduced sensors of CA-Stream. Results are mean ± standard deviation over three runs. Lower is better; **bold** and underlined denote the best and second-best methods within each period, sensor group, adaptation budget, and horizon.

Budget	Sensor group	FNO				EVOLVE				STFO-Large			
		$h=3$	$h=6$	$h=12$	*Avg.*	$h=3$	$h=6$	$h=12$	*Avg.*	$h=3$	$h=6$	$h=12$	*Avg.*
P1: 480 existing sensors; 211 newly introduced sensors													
Zero-shot	Existing	42.26±0.87	43.55±1.75	48.20±0.23	44.36±0.94	44.15±0.86	45.18±0.99	48.58±1.25	45.65±1.04	**38.11**±0.33	**36.63**±0.97	**39.66**±2.83	**38.21**±0.96
	Newly introduced	60.70±5.50	60.78±2.52	**61.24**±4.64	61.21±4.10	63.85±1.74	65.25±1.85	69.52±1.85	65.91±1.97	**59.93**±2.90	**57.73**±2.85	62.96±2.55	**60.42**±1.81
Full	Existing	15.85±0.07	18.06±0.27	21.92±0.04	18.16±0.08	16.44±0.12	18.61±0.16	24.08±0.25	19.18±0.16	**14.13**±0.13	**15.18**±0.12	**16.87**±0.48	**15.18**±0.13
	Newly introduced	16.49±0.15	18.84±0.76	22.09±0.09	18.70±0.22	17.49±0.27	19.73±0.15	26.02±0.11	20.48±0.15	**14.96**±0.24	**16.08**±0.18	**17.76**±0.84	**16.04**±0.21
P2: 691 existing sensors; 484 newly introduced sensors													
Zero-shot	Existing	20.93±0.36	24.26±0.57	30.75±0.72	24.72±0.43	19.75±0.10	**21.64**±0.25	26.31±0.44	22.15±0.22	**19.54**±0.55	21.72±0.60	**25.96**±0.12	**21.97**±0.18
	Newly introduced	21.69±0.30	25.25±0.35	33.09±0.92	26.00±0.46	**18.87**±0.11	**20.91**±0.20	25.80±0.38	21.41±0.18	19.15±0.39	21.05±0.60	**24.68**±0.17	**21.27**±0.17
Full	Existing	15.48±0.22	17.28±0.18	20.82±0.14	17.55±0.18	16.28±0.47	18.34±0.53	23.80±0.87	18.95±0.59	**14.10**±0.05	**15.03**±0.12	**17.04**±0.18	**15.25**±0.08
	Newly introduced	15.30±0.12	16.92±0.15	19.92±0.17	17.12±0.06	16.26±0.36	18.01±0.37	22.78±0.63	18.54±0.44	**13.94**±0.05	**14.68**±0.12	**16.38**±0.19	**14.90**±0.09
P3: 1,175 existing sensors; 523 newly introduced sensors													
Zero-shot	Existing	21.46±0.56	22.52±0.05	**25.48**±0.23	22.77±0.23	**20.91**±1.36	**22.23**±1.12	**25.48**±1.12	**22.57**±1.20	22.03±0.51	23.71±0.12	25.83±0.49	23.38±0.15
	Newly introduced	26.48±0.70	28.51±0.54	34.69±1.43	29.16±0.73	**22.30**±1.81	**23.59**±1.42	**26.79**±1.33	**23.94**±1.53	26.06±0.66	28.32±0.21	31.99±1.16	28.15±0.49
Full	Existing	14.72±0.35	16.50±0.22	19.36±0.10	16.55±0.08	15.14±0.06	17.12±0.07	21.84±0.09	17.58±0.06	**13.72**±0.07	**14.96**±0.03	**17.00**±0.13	**15.04**±0.05
	Newly introduced	17.45±0.33	19.81±0.43	23.91±0.16	19.99±0.09	17.15±0.07	19.14±0.12	23.77±0.11	19.62±0.11	**16.17**±0.17	**17.51**±0.08	**20.07**±0.19	**17.71**±0.09

Table 11: Period-wise RMSE on existing and newly introduced sensors of AIR-Stream. Results are mean ± standard deviation over three runs. Lower is better; **bold** and underlined denote the best and second-best methods within each period, sensor group, adaptation budget, and horizon.

Budget	Sensor group	FNO				EVOLVE				STFO-Large			
		$h=3$	$h=6$	$h=12$	*Avg.*	$h=3$	$h=6$	$h=12$	*Avg.*	$h=3$	$h=6$	$h=12$	*Avg.*
2017: 1,087 existing sensors; 67 newly introduced sensors													
Zero-shot	Existing	28.86±0.45	36.61±0.22	42.36±0.42	34.61±0.27	28.81±0.08	37.89±0.06	44.14±0.02	35.66±0.03	**27.91**±0.20	**35.78**±0.37	**41.76**±0.62	**33.87**±0.38
	Newly introduced	28.55±0.53	36.56±0.35	43.49±0.79	34.83±0.49	28.96±0.10	38.85±0.04	46.67±0.03	36.78±0.03	**27.43**±0.17	**35.60**±0.38	**42.80**±0.60	**33.96**±0.39
Full	Existing	27.92±0.17	35.58±0.36	41.48±0.41	33.68±0.30	28.84±0.36	37.97±0.24	44.22±0.17	35.73±0.28	**27.47**±0.07	**35.00**±0.07	**40.94**±0.09	**33.21**±0.06
	Newly introduced	27.39±0.21	35.58±0.55	43.03±0.77	33.98±0.46	29.00±0.44	38.94±0.29	46.92±0.22	36.91±0.35	**26.92**±0.12	**34.63**±0.03	**41.97**±0.08	**33.20**±0.02
2018: 1,154 existing sensors; 39 newly introduced sensors													
Zero-shot	Existing	25.25±0.07	32.35±0.16	**38.18**±0.44	30.79±0.12	26.78±0.63	34.61±0.32	40.16±0.10	32.74±0.41	**25.10**±0.04	**32.25**±0.18	38.22±0.33	**30.72**±0.18
	Newly introduced	25.45±0.10	33.62±0.29	40.17±0.45	31.98±0.21	27.42±0.79	35.72±0.43	42.08±0.14	33.94±0.52	**24.99**±0.13	**32.47**±0.25	**38.96**±0.29	**30.99**±0.18
Full	Existing	25.04±0.07	32.23±0.18	37.82±0.21	30.53±0.14	26.32±0.15	35.12±0.42	41.23±0.57	33.03±0.33	**24.85**±0.03	**32.14**±0.09	**37.66**±0.17	**30.37**±0.08
	Newly introduced	25.18±0.05	33.35±0.25	39.69±0.47	31.57±0.23	26.67±0.20	36.08±0.35	42.90±0.45	33.98±0.25	**24.79**±0.04	**32.42**±0.16	**38.37**±0.17	**30.67**±0.13
2019: 1,193 existing sensors; 9 newly introduced sensors													
Zero-shot	Existing	24.10±0.36	30.26±0.29	35.18±0.08	28.83±0.14	24.98±0.54	32.80±0.30	37.93±0.18	30.86±0.37	**23.90**±0.12	**30.17**±0.19	**35.15**±0.18	**28.74**±0.14
	Newly introduced	35.13±0.60	**42.38**±0.55	**47.57**±0.55	**40.20**±0.40	35.53±0.57	44.55±0.16	49.36±0.24	41.79±0.25	**34.98**±0.09	42.60±0.34	47.85±0.43	40.39±0.30
Full	Existing	**23.39**±0.14	**29.97**±0.15	35.14±0.25	**28.41**±0.17	24.39±0.09	32.32±0.02	37.71±0.09	30.35±0.07	23.49±0.18	30.03±0.24	**35.01**±0.30	**28.41**±0.20
	Newly introduced	**34.51**±0.16	**41.90**±0.25	**46.82**±0.22	**39.53**±0.17	35.39±0.22	44.40±0.20	49.51±0.32	41.59±0.24	34.82±0.12	42.70±0.13	48.09±0.24	40.31±0.05

Table 12: Period-wise MAPE on existing and newly introduced sensors of AIR-Stream. Results are mean ± standard deviation over three runs. Lower is better; **bold** and underlined denote the best and second-best methods within each period, sensor group, adaptation budget, and horizon.

Budget	Sensor group	FNO				EVOLVE				STFO-Large			
		$h=3$	$h=6$	$h=12$	*Avg.*	$h=3$	$h=6$	$h=12$	*Avg.*	$h=3$	$h=6$	$h=12$	*Avg.*
2017: 1,087 existing sensors; 67 newly introduced sensors													
Zero-shot	Existing	21.96±0.48	29.41±0.13	**34.35**±0.43	27.31±0.18	21.46±0.19	30.05±0.18	35.61±0.06	28.01±0.14	**21.33**±0.21	**28.61**±0.26	35.01±0.98	**27.16**±0.43
	Newly introduced	21.61±0.40	29.69±0.35	**36.60**±0.70	27.95±0.29	21.02±0.21	30.08±0.25	37.28±0.10	28.35±0.18	**20.94**±0.30	**28.78**±0.53	36.84±1.26	**27.62**±0.66
Full	Existing	21.28±0.30	29.12±0.51	34.38±0.64	27.28±0.45	21.73±0.37	30.55±0.35	36.39±0.28	28.51±0.33	**21.04**±0.18	**28.32**±0.49	**33.81**±0.39	**26.64**±0.34
	Newly introduced	20.87±0.28	29.07±0.37	35.68±0.63	27.50±0.38	21.29±0.31	30.43±0.32	37.99±0.30	28.77±0.30	**20.47**±0.14	**28.07**±0.41	**34.75**±0.40	**26.62**±0.31
2018: 1,154 existing sensors; 39 newly introduced sensors													
Zero-shot	Existing	22.39±0.47	31.26±0.68	37.48±0.69	29.52±0.62	22.92±0.46	32.01±0.38	38.54±0.28	30.12±0.36	**22.21**±0.29	**30.36**±0.95	**36.19**±0.36	**28.50**±0.57
	Newly introduced	25.34±0.68	36.75±1.17	44.29±1.17	34.57±1.07	24.83±0.53	34.64±0.46	41.59±0.36	32.63±0.45	**24.13**±0.34	**33.44**±1.17	**39.67**±0.66	**31.22**±0.68
Full	Existing	22.20±0.29	30.39±0.30	37.10±1.05	28.79±0.51	21.86±0.28	30.89±0.47	37.75±0.64	29.11±0.44	**21.58**±0.06	**29.75**±0.21	**36.18**±0.14	**28.07**±0.15
	Newly introduced	25.47±0.48	36.38±0.74	44.81±2.19	34.32±1.08	23.67±0.30	33.48±0.44	40.85±0.68	31.56±0.45	**23.66**±0.02	**32.99**±0.30	**39.74**±0.11	**30.96**±0.17
2019: 1,193 existing sensors; 9 newly introduced sensors													
Zero-shot	Existing	25.81±0.33	33.89±0.42	**40.41**±1.24	32.29±0.58	**24.68**±0.85	34.11±0.96	41.34±1.25	32.28±0.98	25.30±0.12	**33.60**±0.22	40.67±0.25	**32.15**±0.09
	Newly introduced	25.75±0.42	**33.36**±0.95	**38.64**±0.76	**31.38**±0.62	**25.00**±0.96	33.65±1.13	39.26±1.18	31.54±1.07	25.54±0.11	33.56±0.03	40.18±0.67	31.95±0.24
Full	Existing	21.68±0.43	29.76±1.06	35.74±0.96	27.95±0.76	21.84±0.05	31.18±0.08	38.23±0.21	29.31±0.08	**21.33**±0.15	**29.15**±0.33	**35.61**±0.79	**27.54**±0.36
	Newly introduced	23.08±0.33	30.80±0.75	**36.01**±1.29	28.72±0.64	22.73±0.10	31.49±0.05	37.54±0.09	29.45±0.03	**22.44**±0.10	**30.16**±0.53	36.56±1.22	**28.51**±0.63

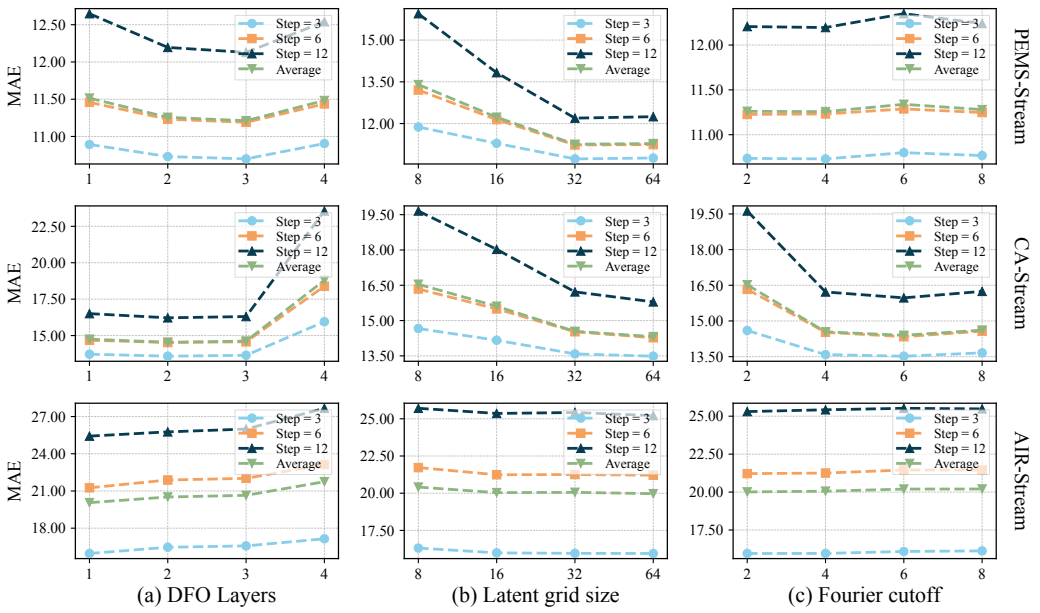

Figure 8: Hyperparameter sensitivity of **STFO**-Large across three streams. Columns show variations in DFO depth, latent grid size, and Fourier cutoff. Curves report MAE at horizons 3, 6, and 12, together with average MAE.

I HYPERPARAMETER SENSITIVITY

We address RQ6 by examining sensitivity to DFO depth, latent grid size, and the Fourier cutoff M. All experiments use the Large configuration and follow the protocol in Section 4.1. Figure 8 illustrates how these choices affect forecasting accuracy across streams and prediction horizons.

❶ ***DFO depth.*** The lowest average MAE is achieved with three layers on PEMS-Stream, two on CA-Stream, and one on AIR-Stream. Increasing depth up to three layers improves performance on PEMS-Stream, while CA-Stream shows little variation between one and three layers. Errors on AIR-Stream increase with depth. Using four layers worsens performance on all three streams relative to their best settings, with a particularly pronounced increase at horizon 12 on CA-Stream. Thus, increasing operator depth does not consistently improve forecasting.

❷ ***Latent grid size.*** Increasing the number of grid points per axis from 8 to 32 substantially reduces errors on both traffic streams, especially at horizon 12. PEMS-Stream shows little further change at 64, while CA-Stream continues to improve, with smaller gains than at coarser resolutions. AIR-Stream benefits mainly from increasing the grid size from 8 to 16, with comparatively small changes thereafter. These results indicate greater sensitivity to spatial resolution on the traffic streams, with diminishing gains as the grid becomes finer.

❸ ***Fourier cutoff.*** PEMS-Stream, CA-Stream, and AIR-Stream achieve their lowest average MAE at $M = 4$, $M = 6$, and $M = 2$, respectively. CA-Stream is the most sensitive: increasing M from 2 to 4 sharply reduces errors, $M = 6$ yields a further modest improvement, and $M = 8$ slightly increases errors. PEMS-Stream varies little across the tested settings, while AIR-Stream shows a slight increase in error as M increases. The benefit of a wider spectral bandwidth therefore depends on the stream.

Overall, **STFO** performs well with shallow operator stacks, while the preferred grid resolution and spectral bandwidth vary across streams. Increasing these hyperparameters beyond suitable settings yields limited gains or degrades performance, supporting the use of compact field operators with spatial and spectral resolutions suited to each forecasting task.

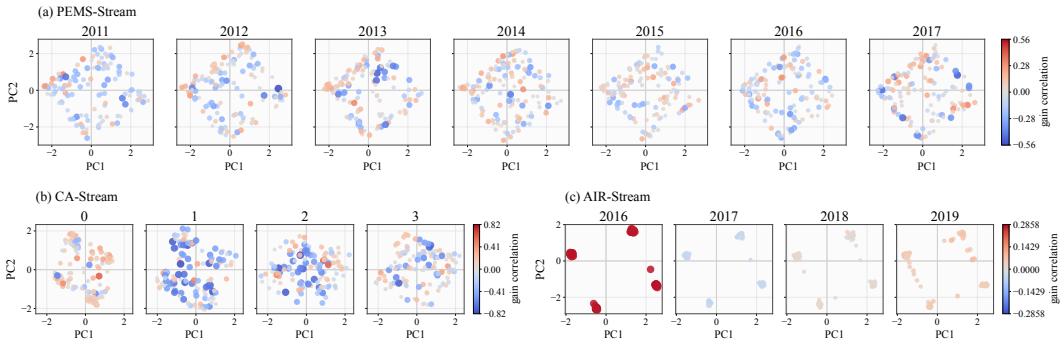

Figure 9: FiLM channel projections across three streams and periods. Channels are projected by PCA from concatenated γ_j–SRE and δ_j–SRE correlations using a shared basis within each stream. Color and size respectively encode the signed and absolute correlations of modulation strength $|\gamma_j|+|\delta_j|$ with forecasting gain $\mathrm{MAE}_{\mathrm{local}} - \mathrm{MAE}_{\mathrm{full}}$ at horizon 6.

J Additional Visual Diagnostics

Figure 9 relates channel modulation to forecasting gains. Each point represents a channel, projected by PCA from its concatenated γ_j–SRE and δ_j–SRE correlations, using a shared basis across periods within each stream. Color and size encode the signed and absolute correlations between modulation strength $|\gamma_j| + |\delta_j|$ and forecasting gain $\mathrm{MAE}_{\mathrm{local}} - \mathrm{MAE}_{\mathrm{full}}$ at horizon 6. PEMS-Stream and CA-Stream show dispersed channel configurations with mixed gain correlations, whereas AIR-Stream exhibits tighter clusters with similar correlations within each period. These patterns suggest that the association between channel modulation and forecasting gains varies across datasets and periods.